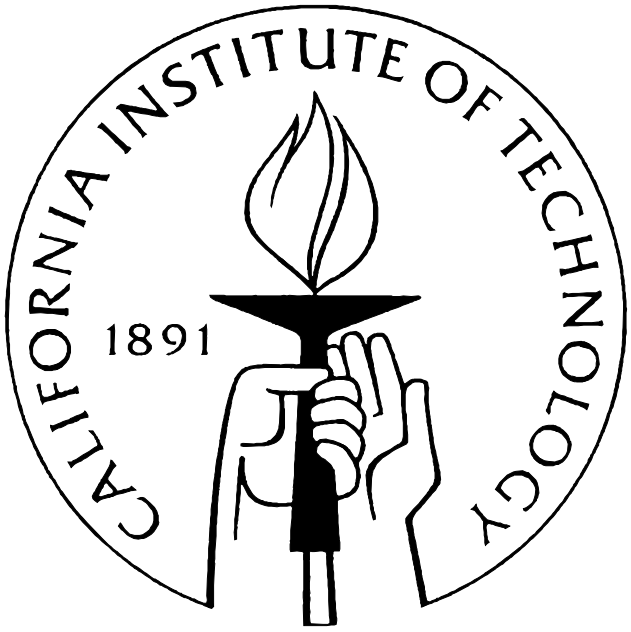

Quantum Information Theory

Classical Communication over Quantum Channels

John A. Cortese

Physics Department

California Institute of Technology

Pasadena, California

2003

Quantum Information Theory
Classical Communication over Quantum Channels

Thesis by
John A. Cortese

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

2003
(Defended October 3, 2003)

Acknowledgments

Committee Members

The author would like to thank the members of his candidacy and thesis defense committees for their time and assistance: Professor John Preskill, Professor Jeff Kimble, Professor Hideo Mabuchi, and Professor Kip Thorne.

Chapter I Acknowledgments

The author would like to thank Professor Benjamin Schumacher for countless interesting conversations regarding channel capacity and channel additivity. The author would also like to thank Dr. Patrick Hayden, Dr. Eric Rains, and Dr. David Bacon for comments and discussions during the development of this work.

Chapter II Acknowledgments

The author would like to thank Patrick Hayden, Charlene Ahn, Sumit Daftuar and John Preskill for helpful comments on a draft of this chapter. The author would also like to thank Beth Ruskai for many interesting conversations on the classical channel capacity of qubit channels.

Chapter III Acknowledgments

The author would like to thank the Institute for Scientific Interchange Foundation, Turin, Italy and Elsag, a Finmeccanica company, for sponsoring seminars in Torino during which some of the work in this chapter was done. In addition, I would like to thank Chris Fuchs, John Smolin and Andrew Doherty for useful discussions.

Chapter IV Acknowledgments

The author would like to thank Andrew Doherty for useful discussions.

Institute for Quantum Information

The author would like to thank members of the Institute for Quantum Information for many stimulating conversations. Specifically, I would like to thank Dave Bacon, Dave Beckman, Andrew Doherty, Christopher Fuchs, Daniel Gottesman, Patrick Hayden, and Andrew Landahl for interesting discussions and helpful commentaries on my research.

Space Radiation Laboratory

The author would like to thank the Caltech Space Radiation Laboratory for providing office space and computer equipment during my time in the Caltech Physics department. The SRL computer staff, Minerva Calderon and Kimberly Rubal, provided computer equipment and support, as well as friendship. I have been especially lucky to have great officemates over the years: Derrick Bass, Brad Cenko, Hubert Chen, Megan Eckart, Allan Labrador, Peter Mao and Brian Matthews ! Thanks for everything - the bookshelve space, the cold temps, and much more !

My Family

The author would like to thank his family: brother David, sister Susan, and especially my mother, for support and assistance over the past several years.

Abstract

This thesis studies classical communication over quantum channels. Chapter 1 describes an algebraic technique which extends several previously known qubit channel capacity results to the qudit quantum channel case. Chapter 2 derives a formula for the relative entropy function of two qubit density matrices in terms of their Bloch vectors. The application of the Bloch vector relative entropy formula to the determination of Holevo-Schumacher-Westmoreland (HSW) capacities for qubit quantum channels is discussed. Chapter 3 outlines several numerical simulation results which support theoretical conclusions and conjectures discussed in Chapters 1 and 2. Chapter 4 closes the thesis with comments, examples and discussion on the additivity of Holevo χ and the HSW channel capacity.

Contents

List of Figures

List of Tables

Chapter 1

Classical Communication over Quantum Channels An Algebraic Analysis

I Abstract

In this chapter, we present a broad introduction to classical communication over quantum channels. The main analytic result is a proof that for a special class of *qudit* unital channels, the Holevo-Schumacher-Westmoreland channel capacity is $\mathcal{C} = \log_2(d) - min_\rho \mathcal{S}(\mathcal{E}(\rho))$, where d is the dimension of the qudit. The result is extended to products of the same class of unital qudit channels. Channel capacity additivity is defined, and the relationship between the minimum entropy at the channel output and channel additivity is outlined for the class of channels of interest in this chapter. The connection between the minimum von Neumann entropy at the channel output and the transmission rate of classical information over quantum channels is extended beyond the previously known qubit result[1].

II Introduction

This chapter describes the communication of classical information (bits) over quantum channels. The main topic is the analysis of various classes of quantum channels using a recently introduced algebraic technique. This technique allows results derived for spin $\frac{1}{2}$ quantum channels to be extended to higher spin dimensional systems.

The transmission of classical information over quantum channels that we shall consider is

[1] A preprint of this work can be found on the Los Alamos National Laboratory preprint server[1].

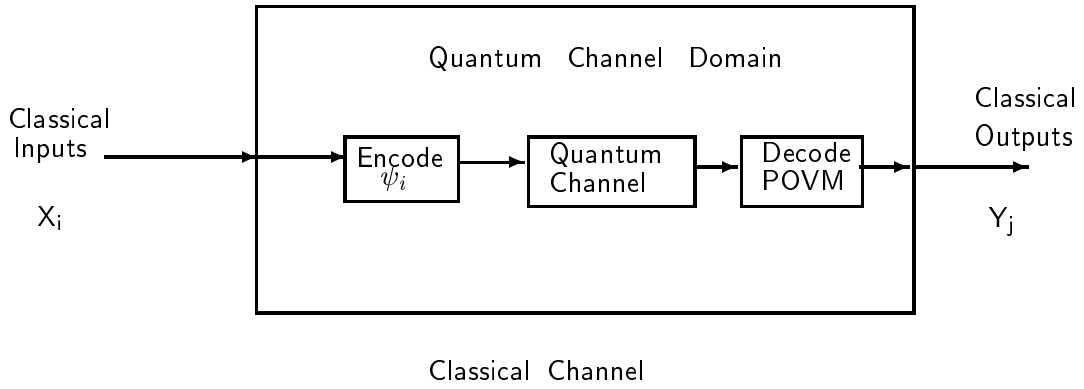

Figure 1.1: Transmission of classical information through a quantum channel.

with no prior quantum entanglement between the sender (Alice) and the recipient (Bob). In such a scenario, classical information is encoded into a set of quantum states ψ_i. These states are transmitted over a quantum channel, during which the signals are perturbed. A receiver at the channel output measures the perturbed quantum states. The resulting measurement outcomes represent the extraction of classical information from the channel output quantum states. The complete scenario is shown above.

The chapter consists of three parts. The first part provides background material. The second part introduces an algebraic formulation of the channel capacity problem, and describes those channels for which an algebraic analysis is applicable. The third part discusses the issue of channel additivity. Channel additivity is currently an active area of research in Quantum Information Theory. Additivity was the motivation behind the development of the algebraic channel capacity analysis technique, and the connections between the two problems are outlined. Open questions related to the additivity of quantum channel capacities are discussed.

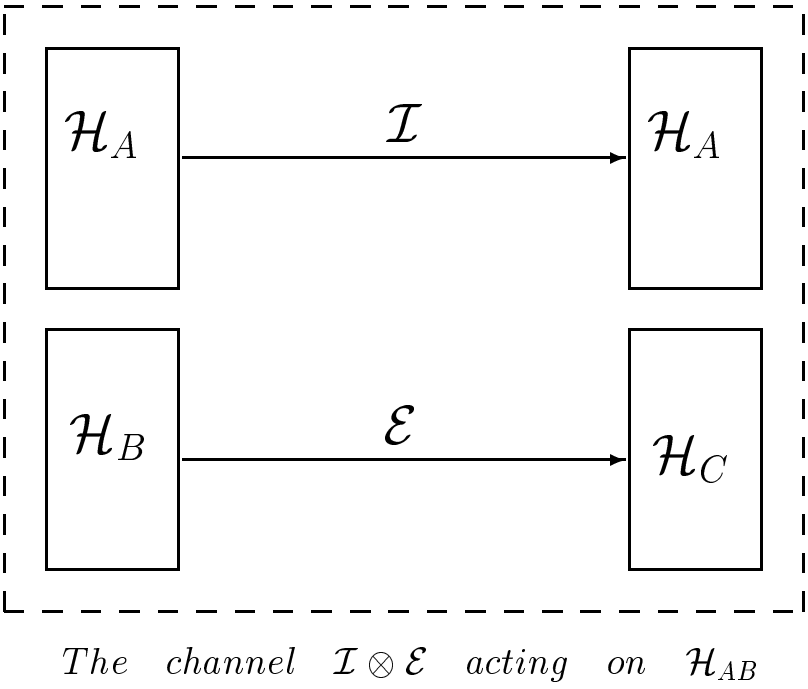

Figure 1.2: Complete Positivity of the quantum channel $\mathcal{E}$.

III Quantum Channels

A quantum channel is a physically realizable mapping of valid density operators to valid density operators. The density operators represent the quantum state of a particle. In this work, we consider only finite dimensional Hilbert spaces. Accordingly, the quantum states are frequently thought of as describing the spin state of a particle. The formal mathematical definition of a physically realizable quantum channel has been shown to be a linear, trace preserving, completely positive map[2]. A positive map take matrices with non-negative eigenvalues to matrices with non-negative eigenvalues. The "complete" qualifier derives from the fact that the system of interest may be considered to be a subsystem of a larger quantum system. Let $\mathcal{E}_B : \mathcal{H}_B \longrightarrow \mathcal{H}_C$. The map $\mathcal{E}_B$ is completely positive if, and only if, for all possible Hilbert spaces $\mathcal{H}_A$, the map $\mathcal{I}_A \otimes \mathcal{E}_B$ acting on the Hilbert space $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$, is a positive map[2]. Here $\mathcal{I}_A$ is the identity map on the Hilbert space $\mathcal{H}_A$, so $\mathcal{I}_A : \mathcal{H}_A \longrightarrow \mathcal{H}_A$, and $\mathcal{I}_A \otimes \mathcal{E}_B : \mathcal{H}_A \otimes \mathcal{H}_B \longrightarrow \mathcal{H}_A \otimes \mathcal{H}_C$.

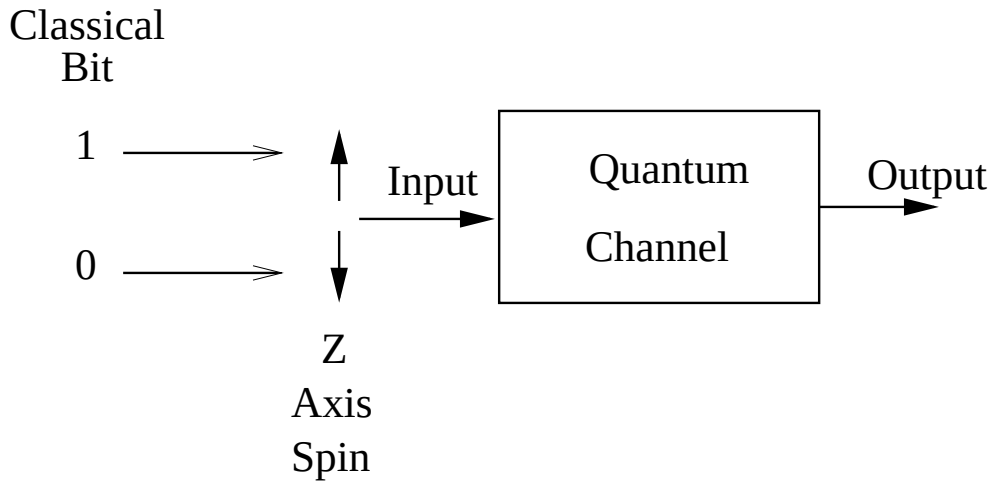

Figure 1.3: Encoding classical bits into the spin of an electron.

Many maps are positive but not completely positive. An example is the partial transpose map $\mathcal{E}(\rho) = \rho^T$, where T denotes the non-conjugated matrix transposition operation on the density matrix ρ[2]. Mathematical techniques have been developed to determine whether a map is completely positive[3, 4], and one such technique is used below. The linearity and trace preserving nature of a quantum channel are as usually defined. For a channel map $\mathcal{E}$, complex coefficients α and β, and density matrices ρ and ϕ, linearity implies

$$\mathcal{E}(\alpha\rho + \beta\phi) = \alpha\mathcal{E}(\rho) + \beta\mathcal{E}(\phi),$$

while trace preserving is defined by the condition that $\forall\rho$, $Trace[\mathcal{E}(\rho)] = Trace[\rho]$.

The transmission of classical data over a quantum channel is implemented in three steps.

1) Source Coding: Classical data (bits) are encoded into the quantum state of a particle. The generic example is the encoding of classical binary data into the Z axis spin component of a spin $\frac{1}{2}$ particle such as an electron. In the nomenclature of quantum information theory, a two level quantum system such as the spin degrees of freedom of an electron, is called a *qubit*. A general d - level spin system is called a *qudit*. A classical bit to qubit encoding is shown below.

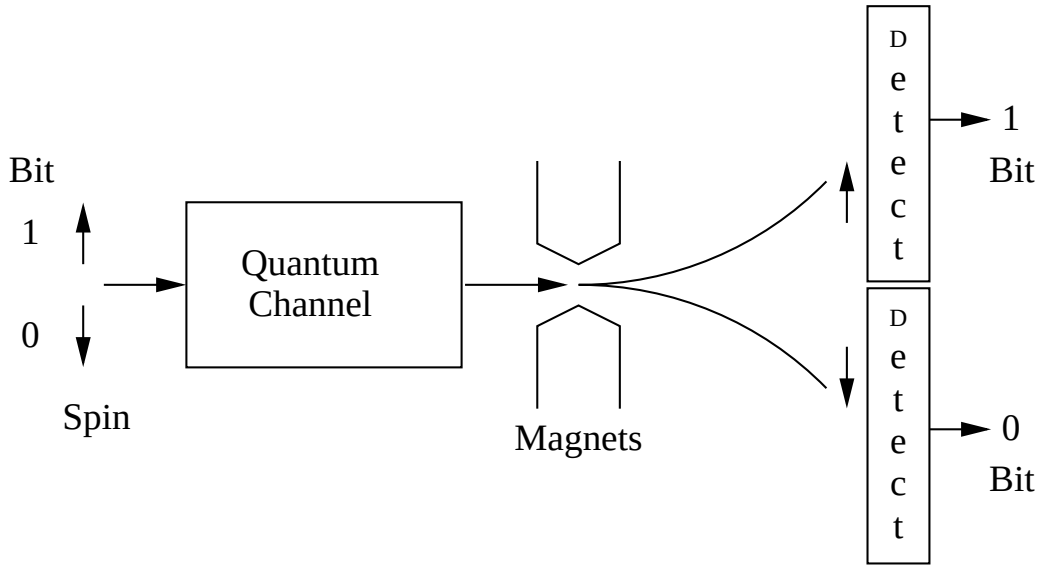

Figure 1.4: Stern - Gerlach detection of electron spin.

2) Channel Evolution: Once the classical data has been encoded into a particle, the density operator representing the quantum state of the particle is evolved via the quantum channel map $\mathcal{E}$. Trivial examples are the identity map $\mathcal{E}(\rho) \longrightarrow \rho$, and the projection map onto the maximally mixed state[2] $\mathcal{E}(\rho) \longrightarrow \frac{1}{d} I_d$.

3) Measurement: Measurement serves to extract the classical information encoded in the quantum particle back into the classical world. For example, in the bit $\rightarrow$ qubit encoding example in 1) above, we measure the Z axis spin component of the electron. If the channel was indeed the identity, we expect to recover the classically encoded data with zero probability of error. Typically the channel perturbs the incoming quantum state, and as a result, the recovery measurement will have a non-zero probability of error.

As a physical example of the entire process, consider the system below.

[2]We write I_d to denote the d by d identity matrix.

The source coding maps classical binary data onto the Z axis spin orientation of an electron. The preparation scheme can be implemented as follows. Imagine having a pool of spin up electrons, and a pool of spin down electrons. Depending on the classical bit, choose an electron from one of these input bins, and transmit that electron through the channel.

The channel itself can be thought of as physically moving the electron from one location to another, for example from the earth to the moon. Along the way, the electron may encounter magnetic fields, which would cause the spin axis of the electron to precess, thereby disturbing the spin direction, and perturbing the quantum state of the electron. Classically, we interpret this spin precession as "noise". It is assumed that the magnetic fields the electron encounters along it's journey are unknown, as otherwise we could compensate for the precession at the channel output before a measurement of the spin is implemented.

The spin measurement can be done via the Stern-Gerlach effect[5, 6, 7]. The detectors in Figure 1.4 indicate the impact of an electron. Which detector "goes off" for each electron output from the channel indicates the Z axis spin orientation - up versus down. (Think of the detectors as geiger counters.)

If the quantum channel above is noiseless (i.e.: no magnetic fields present and no other sources of perturbations for transiting signalling states), we obtain, using the encoding and decoding schemes in Figure 1.4, a channel capacity of one bit of classical information for each electron passing through the channel. From classical communication theory, one would expect that for a noiseless quantum channel, by employing more sophisticated encoding and decoding schemes, the channel capacity would be infinite. Such a line of reasoning is based on the idea that continuous degrees of freedom in noiseless classical systems can encode an infinite amount of binary data. In the case of the electron, the continuous degree of freedom is the direction of the spin axis, which a more sophisticated encoding/decoding could uniformly distribute across the surface of the unit sphere. With no noise present, if this were a classical spin direction, we could recover the spin direction at the channel output with zero probability of error. In quantum mechanics, the information/disturbance tradeoff [8, 9, 10, 11] prevents us from storing an infinite amount of classical information

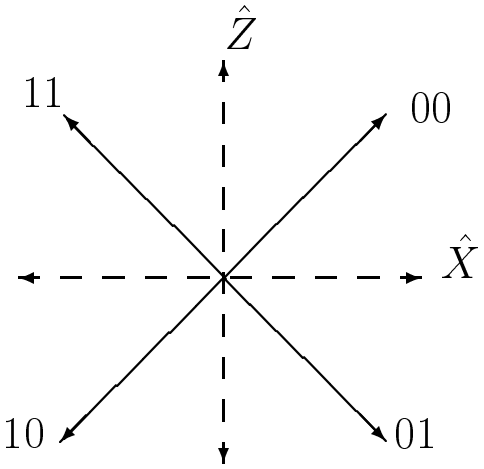

Figure 1.5: Encoding two classical bits into an electron spin.

in continuous quantum degrees of freedom. The limitations from the set of possible quantum measurements we could implement to extract classical information from a quantum state serve to severely reduce the amount of classical information one can store in continuous quantum degrees of freedom. The Holevo theorem on accessible information is a mathematical expression of this limitation[2]. The limitations of quantum measurements in extracting classical information from quantum states is the key issue behind the question of channel capacity additivity to be discussed below.

As an example, consider encoding two classical bits into four equiprobable electron spin orientations in the $\hat{X}$, $\hat{Z}$ plane, as shown in Figure 1.5.

These four states are called a signalling ensemble, as we are trying to communicate using the ensemble $\{ p_i , \psi_i \} = \left\{ \frac{1}{4}, \pm \frac{\hat{X} \pm \hat{Z}}{\sqrt{2}} \right\}$.

Consider the noiseless channel scenario. A classically inspired decoding scheme would first measure the electron spin projection along the $\hat{X}$ direction, followed by a $\hat{Z}$ direction measurement. These two measurements yield two classical bits of information, one bit from each measurement. If the electron spin were a classical phenomenon, this approach would work. However, for a quantum spin, we have the information/disturbance tradeoff. Quantum mechanically, the result of the $\hat{X}$ measurement would yield one classical bit of information, but

would also leave the electron spin in an eigenstate of the $\hat{X}$ operator. The $\hat{Z}$ measurement would return a random binary result, uncorrelated with the original $\hat{Z}$ spin component of the electron. The information context of the second measurement would be zero, and the total information extraction from the quantum state is therefore $1 + 0 = 1$ classical bit. Indeed, the Holevo theorem tells us the maximum amount of classical information which can be extracted from a spin $\frac{1}{2}$ particle such as an electron is one classical bit.

The $\hat{X}$ and $\hat{Z}$ measurements used in the decoding scheme above are orthogonal projective measurements. There are more general types of measurements, known as POVM's (**P**ositive **O**perator **V**alued **M**easurements). POVM's acting on a Hilbert space $\mathcal{B}$ are orthogonal projective measurements on a larger Hilbert space $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$. However, even using POVM's measurements, we would ultimately arrive at the same conclusion: Only one classical bit of information can be extracted from a spin $\frac{1}{2}$ "signalling ensemble" such as shown in Figure 1.5.

Another "trick" we could try which works classically, but fails in the quantum case, is to copy the unknown state at the output of the channel. The original channel output state could undergo an $\hat{X}$ measurement, while the copy could undergo a $\hat{Z}$ measurement. A quantum implementation using a copy of the channel output state, would appear to solve the information/disturbance tradeoff problem. Alas, the No-Cloning theorem states one cannot clone an unknown quantum state, such as the state exiting the channel[2].

The Holevo theorem on accessible information tells us that whatever measurements and/or other "tricks" we may attempt, with *any* input signalling ensemble, we will only be able to extract, at most, one classical bit of information from the quantum state at the output of the channel (for spin $\frac{1}{2}$ particles).

IV Classical Communication over Classical and Quantum Channels

There are many criteria for measuring the quality of the transmission of classical information over a channel, regardless of whether the channel is classical or quantum. In this chapter, we shall focus on the Classical Information Capacity $\mathcal{C}$ of a Quantum Channel $\mathcal{E}$[12, 13, 14].

For classical systems, one usually models the physically realizable transmission dynamics as a discrete memoryless channel (DMC), meaning the channel induced signal perturbation acts in a statistically independent fashion (memoryless) on each message passing through the channel. The DMC model, when combined with the Shannon Entropy Identity in the box below, allows one to conclude the capacity $\mathcal{C}$ of a DMC can be determined from considerations of a single use of the channel. That is, consider an input constellation of signals $\{x_i\}$ with corresponding a priori probabilities $\{p_i\}$, generating the resulting random variable X. An output random variable constellation Y is similarly defined with probabilities $\{q_j\}$, and output signals $\{y_j\}$. In a single use of the channel, one of the x_i is chosen with a priori probability p_i, and sent though the channel. The channel dynamics produce an output signal y_j with a priori probability q_j.

The classical channel capacity $\mathcal{C}$ represents the maximum number of classical bits of information which can be transmitted with one use of the channel, and which can be received at the channel output with arbitrarily small probability of error. The classical channel capacity $\mathcal{C}$ is [15]:

$$\mathcal{C} \;=\; \mathrm{Max}_{\{all\ possible\ x_i\}} \quad H(X) \;-\; H(\,X\,|\,Y\,)$$

Here $H(X)$ is the Shannon entropy for the discrete random variable X, $X \;\equiv\; \{\; p_i \;=\; prob(x_i)\;\}\,,\; i \;=\; 1\,,\cdots,\,N$. The Shannon entropy $H(X)$ is defined as $H(X) \;=\; -\sum_{i=1}^{N}\; p_i\,\log(\,p_i\,)$. For conditional random variables, we denote the probability of the random variable X given Y as $p(X|Y)$. The corresponding conditional Shannon entropy is defined as $H(X|Y) \;=\; -\sum_{i=1}^{N_X}\sum_{j=1}^{N_Y}\; p(x_i\,,\,y_j\,)\;\log[p(\,x_i\,|\,y_j\,)\,]$. Entropy calculations shall be in bits, so $\log_2$ is used.

Key Shannon Entropy Identity leading to classical channel additivity:

For random variables X and Y, $H(X,Y) = H(X) + H(Y|X)$ where $H(X)$, $H(X,Y)$, and $H(Y|X)$ are defined with the formula above, using the probability distributions $p(x)$, $p(x,y)$ and $p(y|x)$, respectively.

To develop intuition about the channel capacity formula, suppose we have $|X|$ linearly independent and equiprobable input signals x_i, and possible output signals y_j, with $|Y| \geq |X|$. If there is no noise in the channel, then $\mathcal{C} = \log_2(|X|)$. Noise in the channel increases the uncertainty in X given the channel output Y, and thus noise increases $H(\,X\,|\,Y)$, thereby decreasing $\mathcal{C}$ for fixed $H(X)$. Geometrically, the presence of random channel noise causes the channel mapping $x_i \rightarrow y_j$ to change from a noiseless one-to-one relationship, to a stochastic map. We say the possible channel mappings of x_i diffuse, occupying a region θ_i instead of a single unique state y_j. As long as the regions θ_i have disjoint support, the receiver can use Y to distinguish which X was sent. In this disjoint support case, $H(\,X\,|\,Y) \approx 0$ and $\mathcal{C} \approx H(X)$. This picture is frequently referred to as sphere packing, since we view the diffused output signals as roughly a sphere around the point in the output space where the signals would have been deposited had the channel introduced no perturbations. The greater the channel noise, the greater the radius of the spheres. If these spheres can be packed into a specified volume without significant overlap, then the decoder can distinguish the input state transmitted by determining which output sphere the decoded signal falls into.

IV.a Sending Classical Information over Quantum Channels

When sending classical information over a quantum channel, we adhere to the same picture. We seek to maximize $H(\,X\,)$ and minimize $H(\,X\,|\,Y\,)$, in order to maximize the channel capacity $\mathcal{C}$. We encode each classical input signal state $\{\,x_i\,\}$ into a corresponding quantum state ψ_i. Sending ψ_i through the channel, the POVM decoder seeks to predict which x_i was originally sent. (See Figure 1.1.) Similar to the classical picture, the quantum channel will diffuse or smear out the density matrix ρ_i corresponding to the quantum state ψ_i as the

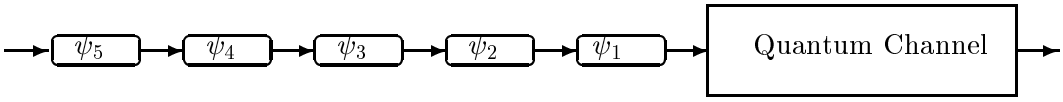

Figure 1.6: Product state inputs ψ_j to a quantum channel.

quantum state passes through the channel. The resulting channel output density matrix $\mathcal{E}(\rho_i)$ will have support over a subspace ϕ_i. As long as all the regions ϕ_i have sufficient disjoint support, the POVM based decoder will be able to distinguish which quantum state ψ_i entered the channel, and hence $H(X|Y) \approx 0$, yielding $\mathcal{C} \approx H(X)$.

For the classical capacity of quantum channels, we encode classical binary data into quantum states. The product state classical capacity for a quantum channel, denoted $\mathcal{C}_1$, maximizes channel throughput by encoding a long block of m classical bits x_i into a long block consisting of the tensor product of n single qudit quantum states ψ_j in a manner which maximizes the (product state) classical channel capacity.

$$\{x_1, x_2, \cdots x_m\} \rightarrow \psi_1 \otimes \psi_2 \otimes \cdots \otimes \psi_n$$

We think of the qudit input states ψ_j as being sequentially transmitted through the qudit channel $\mathcal{E}$.

The tensor product structure of the signalling states $\psi_1 \otimes \psi_2 \otimes \cdots \otimes \psi_n$ does not allow entanglement across the states ψ_k, imitating the memoryless nature of the classical DMC model. The channel perturbation acts independently on each qudit, meaning there is no correlation among any of the "errors" induced by the channel $\mathcal{E}$ on the input qudit states ψ_j.

Since the channel dynamics are known, we can determine the output density matrix for each input density matrix. With the output density matrix for each input signal, we can optimize the measurements (POVMs) performed at the channel output. For example, consider the electron channel in Figure 1.4. The classical data is encoded into the $\pm\hat{Z}$ electron spin

orientation. If the channel were noiseless, implementing a $\pm\hat{Z}$ measurement at the channel output yields a channel capacity of one classical bit per electron passing through the channel. However, if the channel were noiseless, implementing a $\pm\hat{X}$ measurement at the channel output would yield a channel capacity of zero. Implementing the $\hat{X}$ measurement would merely generate random outcomes. To see why, note that

$$|\uparrow\rangle_z \;=\; \frac{|\uparrow\rangle_x \;+\; |\downarrow\rangle_x}{\sqrt{2}} \qquad and \qquad |\downarrow\rangle_z \;=\; \frac{|\uparrow\rangle_x \;-\; |\downarrow\rangle_x}{\sqrt{2}} \;.$$

Here the x and z subscripts denote the basis element along which the spin is up or down. A measurement along the $\hat{X}$ axis of $|\uparrow\rangle_z$ would yield $|\uparrow\rangle_x$ with probability $\frac{1}{2}$ and $|\downarrow\rangle_x$ with probability $\frac{1}{2}$. A measurement along the $\hat{X}$ axis of $|\downarrow\rangle_z$ would also yield $|\uparrow\rangle_x$ with probability $\frac{1}{2}$ and $|\downarrow\rangle_x$ with probability $\frac{1}{2}$. The signalling ensemble in Figure 1.4 calls for $|\uparrow\rangle_z$ and $|\downarrow\rangle_z$ to be input to the channel equiprobably, so measurements at the channel output along the $\hat{X}$ direction would merely generate random outcomes, with no information content.

The main idea in classical communication over quantum channels is that we choose an input ensemble of pure states $\{p_i\,,\,\psi_i\}$ that maximizes the accessible information of the resulting output ensemble. But we must finish the process by choosing a post channel measurement scheme that maximizes the extraction of classical information from the output signalling ensemble. It is important to keep in mind that optimizing the measurement scheme at the channel output is implicit in the capacity calculations to follow.

The Holevo-Schumacher-Westmoreland Channel Capacity Theorem tells us that the classical product state channel capacity using the above encoding scheme is given by the Holevo quantity χ of the output signal ensemble, maximized over a single copy of all possible input signal ensembles $\{p_i\,,\,\rho_i\}$[2].

$$\mathcal{C}_1 \;=\; \mathrm{Max}_{\{all\ possible\ p_i\ and\ \rho_i\}} \quad \chi_{output} \tag{I}$$

$$=\; \mathrm{Max}_{\{all\ possible\ p_i\ and\ \rho_i\}} \quad \mathcal{S}\left(\;\mathcal{E}\left(\sum_i\; p_i\,\rho_i\right)\right) \;-\; \sum_i\; p_i\,\mathcal{S}\left(\;\mathcal{E}\left(\rho_i\right)\;\right)\;.$$

$\mathcal{S}(-)$ is the von Neumann entropy, defined as $\mathcal{S}(\rho) = -Tr\left[\rho \log_2(\rho)\right]$. The symbol $\mathcal{E}(\rho)$ represents the output density matrix obtained from the channel input density matrix ρ.

In implementing the maximization, we need consider only ensembles with at most d^2 elements, where d is the dimension of the qudit input density matrices[16]. (E.g.: For spin $\frac{1}{2}$ particles, $d = 2$.) Since the signalling ensemble contains at most d^2 elements, the channel capacity maximization is of a continuous function over a compact set, and $\mathcal{C}_1$ is attainable[16]. Furthermore, the input signals ρ_i can be chosen to be pure states without affecting the maximization in equation I[16, 12].

In obtaining an optimum signal ensemble satisfying the maximization above, one completely solves the source coding problem (Step 1) and the measurement problem (Step 3) discussed on page 4.

Hereafter we shall call $\mathcal{C}_1$ defined above the Holevo-Schumacher-Westmoreland (HSW) channel capacity[12, 17].

V Additivity of Quantum Channel Capacities

Classical channels are strictly additive, in that classical correlations across sequential input signals to a channel do not increase the classical channel capacity.[3] In determining the product state capacity $\mathcal{C}_1$, we did not allow entanglement, the quantum mechanical version of classical correlation, between any of the quantum signalling states ψ_k in the classical to quantum encoding. Consider allowing entanglement between pairs of successive quantum states ψ_i in the encoding step, as shown below.

$$\{x_1, x_2, \cdots x_m\} \rightarrow \underbrace{\psi_1 \otimes \psi_2}_{entangle} \otimes \overbrace{\psi_3 \otimes \psi_4}^{entangle} \otimes \underbrace{\psi_5 \otimes \psi_6}_{entangle} \otimes \overbrace{\psi_7 \otimes \psi_8}^{entangle} \otimes \cdots \otimes \psi_n$$

$$leading \quad to \quad \{x_1, x_2, \cdots x_m\} \rightarrow \psi_{12} \otimes \psi_{34} \otimes \psi_{56} \otimes \psi_{78} \otimes \cdots \otimes \psi_{n(n-1)} \,,$$

[3] To see why, take a look at Lemma 8.9.2 in [15].

where $\psi_{i(i+1)}$ indicates the states of the i'th and $i+1$'th density matrix inputs to the channel could be entangled with each other, but no other input state. (E.g.: In the electron case, entangle successive pairs of electrons being passed through the channel.) Call the resulting capacity $\mathcal{C}_2$ with this limited degree of entanglement between the ψ_i, along with a corresponding joint two-state measurement scheme across the channel output states to take advantage of the presence of entanglement. Continuing on, we define $\mathcal{C}_N$ as the N-fold entangled input state channel capacity, extending the single fold definition of equation I on page 12. The classical channel capacity of the quantum channel $\mathcal{C}$ is defined as $\mathcal{C} = lim_{N\to\infty}\ \mathcal{C}_N$, thereby allowing unlimited entanglement across all input signalling states. Whether $\mathcal{C} = \mathcal{C}_1$ is the additivity question. One can see immediately that introducing entanglement into the encoding scenario could not decrease the channel capacity below $\mathcal{C}_1$, since product states are a subset of the set of all possible general quantum states, i.e.: a subset of n-qubit states which are allowed to have entanglement.

At first glance, one would think there is substantial benefit to using generic input signalling states which allow entanglement instead of product input signalling states. Consider how the number of free parameters needed to describe n qubit product states scales with n, the number of qubits in the product state Υ.

$$\Upsilon = \begin{bmatrix}\alpha_1\\ \beta_1\end{bmatrix} \bigotimes \begin{bmatrix}\alpha_2\\ \beta_2\end{bmatrix} \bigotimes \begin{bmatrix}\alpha_3\\ \beta_3\end{bmatrix} \bigotimes \ldots \bigotimes \begin{bmatrix}\alpha_n\\ \beta_n\end{bmatrix} \quad (n \quad qubit \quad product \quad state).$$

n qubit product state.

Each qubit above is represented by a two element complex vector. The global phase of each qubit does not matter, and since we lose one degree of freedom from normalization for each qubit, we have two real degrees of freedom per qubit, yielding a total of $P(n) = 2\,n$ degrees of freedom for an n qubit product state.

The situation is much different for a generic n qubit quantum state.

$$\Psi = \mu_1 \left|00\cdots000\right\rangle + \mu_2 \left|00\cdots001\right\rangle + \cdots\cdots + \mu_{2^n-1} \left|11\cdots110\right\rangle + \mu_{2^n} \left|11\cdots111\right\rangle.$$

Generic n qubit state.

Here we have a total of 2^n possible states, and each state has a complex coefficient μ_i. Subtracting a global phase and normalization degrees of freedom, we have a total number of degrees of freedom equal to $G(n) = 2^{n+1} - 2$. Classically, every degree of freedom in a signal is accessible to the receiver and can be used to convey classical information. Intuitively, one might expect to use the many more degrees of freedom for generic states in a redundant fashion, similar to block coding techniques, allowing the receiver to compare several degrees of freedom at the channel output, and determining the true encoded state initially sent. That is, simply repeat the classical 0/1 bit signal in all the degrees of freedom available in a signal. This is a simple repetitive code. Yet, we know from Holevo's work on accessible information that a qubit can carry, at most, one classical bit of information. In the quantum world, the situation is more complicated, limited by the bottleneck of the measurement apparatus. If the quantum receiver could simultaneously and independently look at all of the quantum signal degrees of freedom, one could imbed many more copies of the classical 0/1 information in an n-qubit generic state than an n-qubit product state. Using majority decoding, one would expect greater channel capacity from using n-qubit generic input signalling states than n-qubit product signalling states. Furthermore, the capacity difference should become large very quickly as n increases, given the functional nature of the number of degrees of freedom of n qubit product states versus generic states, namely $2n$ versus $\approx 2^n$.

To better get across the huge difference in degrees of freedom available in entangled states versus product states, we tabulate in Table 1.1 and Figure 1.7 below the (real) degrees of freedom associated with an n qubit product state versus an n qubit generic state, as a function of n.

n (no. of qubits)	2	3	4	5	6	7	8	9	10	11	12
Product	4	6	8	10	12	14	16	18	20	22	24
Generic	6	14	30	62	126	254	510	1022	2046	4094	8190

Table 1.1: Product state degrees of freedom versus generic qubits, for n qubits.

It has been widely conjectured, but not proven[18] that $\mathcal{C} = \mathcal{C}_1$, and that the product state

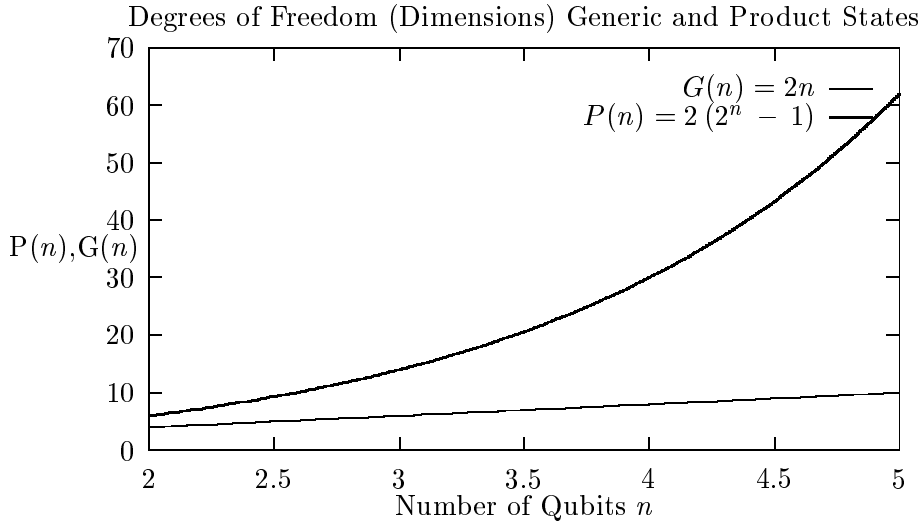

Figure 1.7: Product state degrees of freedom versus generic states, for n qubits.

classical channel capacity $\mathcal{C}_1$ of a quantum channel is the classical capacity $\mathcal{C}$ of the quantum channel. Equality is known for some special channels, such as qubit unital channels[19]. However, whether the classical capacity of a generic quantum channel is additive is unknown. The work in this chapter is a step towards determining a class of qudit channels, valid for all qudit dimensions d, for which the resulting classical channel capacity is additive.

V.a Capacity Additivity and Parallel Quantum Channels

An equivalent picture of channel additivity is the following. Consider two channels acting in parallel. Denote the joint $\mathcal{C}_1$ channel capacity of the tandem channels with a possibly entangled input state ψ_{AB}, and a joint measurement occurring across the two channel outputs, by $\mathcal{C}_{AB}$. The two channels, when used individually, without possibly entangled input states, and no joint output measurement, have $\mathcal{C}_1$ channel capacities $\mathcal{C}_A$ and $\mathcal{C}_B$. The dotted box is intended to indicate channels A and B are tensored together to form a single "super" channel AB. The channel $\mathcal{E}_A : \mathcal{H}_A \longrightarrow \mathcal{H}_A$, and the channel $\mathcal{E}_B : \mathcal{H}_B \longrightarrow \mathcal{H}_B$, while the "super" channel $\mathcal{E}_{AB} : \mathcal{H}_{AB} \longrightarrow \mathcal{H}_{AB}$, where $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$.

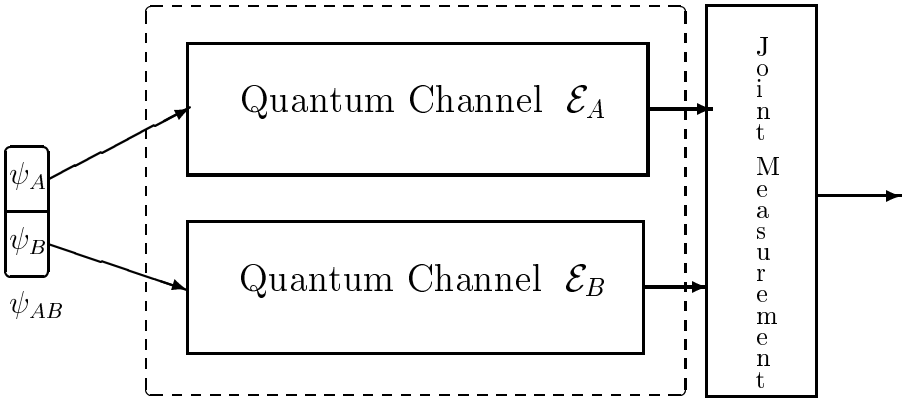

Figure 1.8: Parallel channel view of capacity additivity.

Here $\mathcal{C}_A$ is the HSW channel capacity of $\mathcal{E}_A$, $\mathcal{C}_B$ is the HSW channel capacity of $\mathcal{E}_B$, and $\mathcal{C}_{AB}$ is the HSW channel capacity of $\mathcal{E}_{AB}$. If $\mathcal{C}_{AB} > \mathcal{C}_A + \mathcal{C}_B$, then channel superadditivity holds. That is, entanglement across the input states ψ_{AB} increases the channel capacity, and the quantum channels do not have strictly additive capacities, meaning $\mathcal{C}_{AB} \neq \mathcal{C}_A + \mathcal{C}_B$. It should be noted that the equivalence of these two views of superadditivity, serial entangled inputs versus parallel entangled inputs, is akin to the concept of random process ergodicity in classical communication theory[20, 21].

Reconciling the serial input picture of channel capacity with the parallel channel view, we have $\mathcal{C}_2 = \frac{1}{2}\mathcal{C}_{AB}$, if $\mathcal{E}_A$ and $\mathcal{E}_B$ are the same channel. The N-fold extension is $\mathcal{C}_N = \frac{1}{N}\mathcal{C}_{\mathcal{E}^{\otimes N}}$ and $\mathcal{C} = lim_{N\to\infty}\,\mathcal{C}_N$.

VI Optimal Signalling Ensembles

In equation I, we define $\mathcal{C}_1$, the Holevo-Schumacher-Westmoreland (HSW) channel capacity. An input ensemble $\{p_i\,,\,\rho_i\}$ that achieves $\mathcal{C}_1$ is called an optimal input ensemble. There may be several different optimal input ensembles which achieve the optimum HSW channel capacity $\mathcal{C}_1$. However, it will be shown that the average channel output state of an optimal ensemble is a unique state for all optimal ensembles for that channel. That is, given a set

of optimal input ensembles

$$\left\{p_i^{(1)}, \rho_i^{(1)}\right\} \ , \ \left\{p_i^{(2)}, \rho_i^{(2)}\right\} \ , \ \cdots \ , \ \left\{p_i^{(N)}, \rho_i^{(N)}\right\} \ ,$$

all achieving $\mathcal{C}_1$, we define $\widetilde{\Phi}^{(k)} \ = \ \mathcal{E}\left(\sum_i p_i^{(k)} \, \rho_i^{(k)}\right)$. Then we must have $\widetilde{\Phi}^{(1)} \ = \ \widetilde{\Phi}^{(2)} \ = \ \cdots \ = \ \widetilde{\Phi}^{(N)}$.[4]

The main idea of this chapter is the unique nature of the output ensemble average state of an optimal signalling ensemble for a quantum channel $\mathcal{E}$ yields considerable information about the HSW channel capacity $\mathcal{C}_1$ of a channel.

VI.a Relative Entropy and Channel Capacity

An alternate, but equivalent, description of HSW channel capacity can be made using relative entropy[16]. The relative entropy $\mathcal{D}$ of two density matrices, ϱ and ϕ, is defined as [2, 16, 22, 23]:

$$\mathcal{D}(\, \varrho \,\|\, \phi \,) \ = \ Tr \left[\, \varrho \, \log(\, \varrho \,) \ - \ \varrho \, \log(\, \phi \,) \right] \ .$$

Here Tr[-] is the trace operator. Klein's inequality tells us that $\mathcal{D} \ \geq \ 0$, with $\mathcal{D} \ \equiv \ 0$ iff $\varrho \ \equiv \ \phi$ [2]. The logarithms are base 2.

To see how to represent χ in terms of $\mathcal{D}$, recall that an ensemble of channel output states $\{ \, p_k \, , \, \varrho_k \ = \ \mathcal{E}(\varphi_k) \, \}$ is an optimal ensemble if this ensemble achieves $\mathcal{C}_1$. Consider the optimal signalling state ensemble $\{ \, p_k \, , \, \varrho_k \ = \ \mathcal{E}(\varphi_k) \, \}$. Define ϱ as $\sum_k \ p_k \, \varrho_k$. Consider the following sum:

$$\sum_k \ p_k \ \mathcal{D}(\, \varrho_k \,\|\, \varrho \,) \ = \ \sum_k \ \{ p_k \, Tr[\, \varrho_k \ \log(\, \varrho_k \,) \,] \ - \ p_k \, Tr[\, \varrho_k \ \log(\, \varrho \,) \,] \}$$

$$= \quad \sum_k \ \{ p_k \, Tr \left[\, \varrho_k \ \log(\, \varrho_k \,) \, \right] \} \ - \ Tr \left[\ \sum_k \ \{ p_k \ \varrho_k \ \log(\, \varrho \,) \ \} \right]$$

[4] We shall use ρ to denote a density operator at the channel input, and $\widetilde{\rho}$ as the corresponding channel output density operator.

$$= \sum_k \{ p_k \, Tr[\, \varrho_k \, \log(\, \varrho_k \,) \,] \,\} \; - \; Tr \, [\, \varrho \, \log(\varrho) \,] \quad = \quad \mathcal{S}(\, \varrho \,) \; - \; \sum_k \, p_k \, \mathcal{S}(\, \varrho_k \,) \; = \; \chi \; .$$

Thus the HSW channel capacity $\mathcal{C}_1$ can be written as

$$\mathcal{C}_1 \; = \; \mathrm{Max}_{[all\ possible\ \{p_k \, , \, \varphi_k\}]} \quad \sum_k \, p_k \; \mathcal{D} \, (\, \mathcal{E}(\varphi_k) \, || \, \mathcal{E}(\varphi) \,) \; ,$$

where the φ_k are the quantum states input to the channel and $\varphi \; = \; \sum_k \; p_k \, \varphi_k$.

VII The Schumacher-Westmoreland Relative Entropy Lemmas

In 1999, Benjamin Schumacher and Michael Westmoreland published a paper entitled *Optimal Signal Ensembles* [16] that elegantly described the classical (product state) channel capacity of quantum channels in terms of the relative entropy.

Schumacher and Westmoreland proved the following five properties related to optimal ensembles[16].

I) $\mathcal{D}(\, \varrho_k \, \| \, \varrho \,) \; = \; \mathcal{C}_1 \quad \forall \varrho_k$ in the optimal ensemble, and $\varrho \; = \; \sum p_k \, \varrho_k$.

II) $\mathcal{D}(\, \xi \, \| \, \varrho \,) \; \leq \; \mathcal{C}_1$ where $\{ \, p_k \, , \, \varrho_k \; = \; \mathcal{E}(\varphi_k) \, , \; \varrho \; = \; \sum p_k \, \varrho_k \, \}$ is an optimal ensemble, and ξ is *any* permissible channel output density matrix.

III) There exists at least one optimal ensemble $\{ \, p_k \, , \, \varrho_k \; = \; \mathcal{E}(\varphi_k) \, \}$ that achieves $\mathcal{C}_1$.

IV) Let $\mathcal{A}$ be the set of possible channel output states for a channel $\mathcal{E}$ corresponding to pure state inputs. Define $\mathcal{B}$ as the convex hull of the set of states $\mathcal{A}$. Then for $\varrho \; \in \mathcal{A}$ and $\xi \; \in \mathcal{B}$, we have [5]:

$$\mathcal{C}_1 \; = \; \mathrm{Min}_{\, \xi} \qquad \mathrm{Max}_{\, \varrho} \qquad \mathcal{D} \, (\, \varrho \, \| \, \xi \,) \; .$$

[5] This result was originally derived in [24].

V) For every ξ that satisfies the minimization in IV) above, there exists an optimum signalling ensemble $\{ p_k , \rho_k \}$ such that $\xi \equiv \sum_k p_k \rho_k$.

Building upon this work, we shall study quantum channels, adding the following fact to the Schumacher-Westmoreland analysis.

The average output density matrix for any optimal set of signalling states that achieves the maximum classical channel capacity for a quantum channel is unique.

VIII Background Material

VIII.a Invariance of $\mathcal{S}$ and χ under Unitary Operators

Consider any ensemble $\{p_i, \rho_i\}$. Acting on each ρ_i with the same unitary operator U yields a set of valid quantum states $U\rho_i U^\dagger$ and the ensemble $\left\{p_i, U\rho_i U^\dagger\right\}$. Furthermore, each ρ_i has the same eigenvalues as the corresponding $U\rho_i U^\dagger$. Since von Neumann entropy depends only on a density operators eigenvalues, we conclude $\mathcal{S}(\rho_i) = \mathcal{S}\left(U\rho_i U^\dagger\right)$. Furthermore, this implies the Holevo quantity χ of the ensembles $\{p_i, \rho_i\}$ and $\left\{p_i, U\rho_i U^\dagger\right\}$ is equal, since

$$\chi\left(\left\{p_i, U\rho_i U^\dagger\right\}\right) = \mathcal{S}\left(\sum_i p_i U\rho_i U^\dagger\right) - \sum_i p_i \mathcal{S}\left(U\rho_i U^\dagger\right) \tag{II}$$

$$= \mathcal{S}\left(U\left(\sum_i p_i \rho_i\right)U^\dagger\right) - \sum_i p_i \mathcal{S}\left(U\rho_i U^\dagger\right)$$

$$= \mathcal{S}\left(\sum_i p_i \rho_i\right) - \sum_i p_i \mathcal{S}\left(\rho_i\right) = \chi\left(\left\{p_i, \rho_i\right\}\right).$$

VIII.b Uniqueness of the Average Output Ensemble Density Matrix

In this section, we prove that every optimal input ensemble $\{p_i, \rho_i\}$ has the same average output state, $\mathcal{E}(\rho)$, where $\rho = \sum_i p_i\rho_i$. This result will enable us to write the HSW channel

capacity as $\mathcal{C}_1 = \eta - \sum_i p_i \mathcal{E}(\rho_i)$, where η is a constant. To prove this result, we shall show below that if there exists two optimum signalling ensembles, $\{ p_k , \rho_k \}$ and $\{ p'_k , \rho'_k \}$ of channel output states, then the two resulting average density matrices, $\rho = \sum_k p_k \rho_k$ and $\rho' = \sum_k p'_k \rho'_k$ must be equal, thereby proving the uniqueness of the average channel output state of any optimum signalling ensemble.

The approach uses property IV from Section VI.a:

Theorem:

$$\mathcal{C}_1 \quad = \quad Min_\phi \quad Max_\rho \quad D(\rho \| \phi) \, .$$

The maximum is taken over the set $\mathcal{A}$, while the minimum is taken over the set $\mathcal{B}$. Both $\mathcal{A}$ and $\mathcal{B}$ are defined in property IV of Section VI.a. In order to apply the min max formula above for $\mathcal{C}_1$, we need a result regarding the uniqueness of the average output ensemble density matrix $\rho = \sum_k p_k \rho_k$ for different output optimal ensembles $\{ p_k , \rho_k \}$.

Theorem:

The density matrix ϕ which achieves the minimum in the min-max formula for $\mathcal{C}_1$ above is unique.

Proof:

From property V in Section VI.a, we know the optimal ϕ which attains the minimum above must correspond to the average of a set of signal states of an optimum signalling ensemble. We shall prove the uniqueness of ϕ by postulating there are two optimum output signal ensembles, with possibly different average density matrices, σ and ξ. We will then prove that σ must equal ξ, thereby implying ϕ is unique.

Let $\{ \alpha_i , \rho_i \}$ be an optimum output signal ensemble, with probabilities α_i and density matrices ρ_i, where $\alpha_i \geq 0$ and $\sum_i \alpha_i = 1$. Define $\sigma = \sum_i \alpha_i \rho_i$. By property I in Section VI.a, we know that $\mathcal{D}(\rho_i \| \sigma) = \mathcal{C}_1 \quad \forall \, i$.

Consider a second, optimum output signal ensemble $\{ \beta_j , \phi_j \}$ differing in at least one density matrix ρ_i and/or one probability α_i from the optimum output ensemble $\{ \alpha_i , \rho_i \}$. Define $\xi = \sum_j \beta_j \phi_j$. Consider the quantity $\sum_i \alpha_i \mathcal{D}(\rho_i \| \xi)$. Let us apply Donald's equality, which is discussed in appendix A.

$$\sum_i \alpha_i \mathcal{D}(\rho_i \| \xi) = \mathcal{D}(\sigma \| \xi) + \sum_i \alpha_i \mathcal{D}(\rho_i \| \sigma) .$$

Since $\mathcal{D}(\rho_i \| \sigma) = \mathcal{C}_1 \quad \forall i$, and $\sum_i \alpha_i = 1$, we obtain:

$$\sum_i \alpha_i \mathcal{D}(\rho_i \| \xi) = \mathcal{D}(\sigma \| \xi) + \mathcal{C}_1.$$

From property II in Section VI.a, since ξ is the average of a set of optimal signal states $\{ \beta_j , \phi_j \}$, we know that $\mathcal{D}(\rho_i \| \xi) \leq \mathcal{C}_1 \, \forall i$. Thus $\sum_i \alpha_i \mathcal{D}(\rho_i \| \xi) \leq \mathcal{C}_1$. Combining this inequality constraint on $\sum_i \alpha_i \mathcal{D}(\rho_i \| \xi)$ with what we know about $\sum_i \alpha_i \mathcal{D}(\rho_i \| \xi)$ from Donald's equality, we obtain the two relations:

$$\sum_i \alpha_i \mathcal{D}(\rho_i \| \xi) = \mathcal{D}(\sigma \| \xi) + \mathcal{C}_1 \qquad \textit{and} \qquad \sum_i \alpha_i \mathcal{D}(\rho_i \| \xi) \leq \mathcal{C}_1 .$$

From Klein's inequality, we know that $\mathcal{D}(\sigma \| \xi) \geq 0$, with equality iff $\sigma \equiv \xi$. Thus, the only way the equation

$$\sum_i \alpha_i \mathcal{D}(\rho_i \| \xi) = \mathcal{D}(\sigma \| \xi) + \mathcal{C}_1 ,$$

can be satisfied is if we have $\sigma \equiv \xi$, for then $\mathcal{D}(\sigma \| \xi) = 0$ and we have

$$\sum_i \alpha_i \mathcal{D}(\rho_i \| \xi) = \mathcal{D}(\sigma \| \xi) + \mathcal{C}_1 = \mathcal{C}_1 ,$$

and

$$\sum_i \alpha_i \mathcal{D}(\rho_i \| \xi) = \sum_i \alpha_i D(\rho_i \| \sigma) = \sum_i \alpha_i \mathcal{C}_1 = \mathcal{C}_1 .$$

Therefore, only in the case where $\sigma \equiv \xi$ is Donald's equality satisfied. Since σ and ξ were the average output density matrices for two different, but arbitrary optimum signalling ensembles, we conclude the average density matrices of all optimum signalling ensembles

must be equal, thereby implying ϕ is unique.

$\triangle$ - *End of Proof.*

Only generic properties of the relative entropy were used in the above proof of uniqueness. Therefore the uniqueness result holds for any valid quantum channel (i.e.: any Linear, Completely Positive, Trace Preserving map).

IX Channel Capacity of Single Qubit Unital Channels

As an example of the approach we shall be taking, we derive the HSW channel capacity for single qubit unital channels. This result was previously derived in [25] by a different technique.

IX.a The King-Ruskai-Szarek-Werner Qubit Channel Representation

Consider qubit channels, namely $\mathcal{E}(\varphi) = \varrho$, where φ and ϱ are qubit density matrices. Several authors [25, 26] have developed a nice picture of single qubit maps. Recall that single qubit density matrices can be written in the Bloch sphere representation. Let the density matrices ϱ and φ have the respective Bloch sphere representations:

$$\varphi = \frac{1}{2}\left(\mathcal{I} + \vec{\mathcal{W}}_\varphi \bullet \vec{\sigma}\right) \qquad and \qquad \varrho = \frac{1}{2}\left(\mathcal{I} + \vec{\mathcal{W}}_\varrho \bullet \vec{\sigma}\right).$$

The symbol $\vec{\sigma}$ means the vector of 2 by 2 Pauli matrices

$$\vec{\sigma} = \begin{bmatrix} \sigma_x \\ \sigma_y \\ \sigma_z \end{bmatrix} \quad where \quad \sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

The Bloch vectors $\vec{\mathcal{W}}$ are real three dimensional vectors that have magnitude equal to one when representing a pure state density matrix, and magnitude less than one for a mixed

(non-pure) density matrix.

The King-Ruskai et al. qubit channel representation [25] describes the channel as a mapping of input Bloch vectors $\vec{W}$ to output Bloch vectors $\widetilde{\vec{W}}$, as shown below.

$$\begin{bmatrix} 1 \\ \widetilde{W_x} \\ \widetilde{W_y} \\ \widetilde{W_z} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ t_x & \lambda_x & 0 & 0 \\ t_y & 0 & \lambda_y & 0 \\ t_z & 0 & 0 & \lambda_z \end{bmatrix} \begin{bmatrix} 1 \\ W_x \\ W_y \\ W_z \end{bmatrix}$$

All qubit channels have such a $\{ \lambda_k , t_k \}$ representation. The channel representation is unique up to a unitary operation on the input and output Hilbert spaces, and hence requires a special choice of input and output bases. The t_k and λ_k are real parameters which must satisfy certain constraints in order to ensure the matrix above represents a completely positive qubit map. (Please see [25] for more details.)

From the King-Ruskai et al. qubit channel representation, we see that $\widetilde{\mathcal{W}_k} = t_k + \lambda_k \mathcal{W}_k$ or

$$\mathcal{W}_k = \frac{\widetilde{\mathcal{W}_k} - t_k}{\lambda_k} .$$

It has been shown that $\mathcal{C}_1$ can always be achieved using only pure input states[16, 12]. Therefore, all input signalling Bloch vectors obey $\| \vec{\mathcal{W}} \| = 1$. Thus $\| \vec{\mathcal{W}} \|^2 = 1$, and $\| \vec{\mathcal{W}} \|^2 = 1 = \mathcal{W}_x^2 + \mathcal{W}_y^2 + \mathcal{W}_z^2$ implies

$$\left(\frac{\widetilde{\mathcal{W}_x} - t_x}{\lambda_x} \right)^2 + \left(\frac{\widetilde{\mathcal{W}_y} - t_y}{\lambda_y} \right)^2 + \left(\frac{\widetilde{\mathcal{W}_z} - t_z}{\lambda_z} \right)^2 = 1 .$$

The set of possible channel output states we shall be interested in is the set of channel outputs corresponding to pure state channel inputs. This set of states was defined as $\mathcal{A}$ in section 2.2, and is the *surface* of the ellipsoid shown above. The convex hull of the set of states $\mathcal{A}$ is the solid ellipsoid defined as $\widetilde{\vec{\mathcal{W}}}$ such that

$$\left(\frac{\widetilde{\mathcal{W}_x} - t_x}{\lambda_x} \right)^2 + \left(\frac{\widetilde{\mathcal{W}_y} - t_y}{\lambda_y} \right)^2 + \left(\frac{\widetilde{\mathcal{W}_z} - t_z}{\lambda_z} \right)^2 \leq 1 .$$

It was shown in [25] that the action of a single qubit unital channel $\mathcal{E}$ on an input state ρ could be represented as $\widetilde{\rho} \;=\; \mathcal{E}(\rho)$, where ρ has Bloch vector $\begin{bmatrix} w_x \\ w_y \\ w_z \end{bmatrix}$ and $\widetilde{\rho}$ has Bloch vector $\begin{bmatrix} \lambda_x\, w_x \\ \lambda_y\, w_y \\ \lambda_z\, w_z \end{bmatrix}$. Here the $\lambda_k \in [-1, 1]$. Using the unique nature of the average output state of an optimal signalling ensemble, we shall show the HSW channel capacity is $\mathcal{C}_1 \;= 1 - min_\rho\, \mathcal{S}(\mathcal{E}(\rho))$.

IX.b Achievability of Output Ensembles

We say an ensemble $\{q_j\,,\, \phi_j\}$ at the channel output is *achievable* if there exists an input ensemble $\{q_j\,,\, \varphi_j\}$ such that the $\{\varphi_j\}$ are all valid density operators and $\mathcal{E}(\varphi_j) \;=\; \phi_j \;\; \forall j$. Let us recall some properties of the Pauli matrices $\{\sigma_k\}$. The $\{\sigma_k\}$ obey the relations $\sigma_i\sigma_j \;=\; -\,\sigma_j\sigma_i \; for \; i \neq j$ and $\sigma_i\sigma_j \;=\; I_2 \; for \; i = j$. Thus we find $\sigma_i\sigma_j\sigma_i \;=\; -\sigma_j \; for \; i \neq j$ and $\sigma_i\sigma_j\sigma_i \;=\; \sigma_i \; for \; i = j$. The σ_k are Hermitian, so $\sigma_k^2 \;=\; I_2$ implies the σ_k are unitary, yielding $\sigma_k^\dagger \;=\; \sigma_k$.

Let $\{p_i\,,\, \rho_i\}$be an optimal input ensemble with corresponding output ensemble $\{p_i\,,\, \mathcal{E}(\rho_i) \;= \widetilde{\rho}_i\}$. Apply a Pauli operator σ_k to all the density matrices in $\{p_i\,,\, \mathcal{E}(\rho_i) \;=\; \widetilde{\rho}_i\}$, yielding an ensemble $\{p_i\,,\, \sigma_k\,\widetilde{\rho}_i\,\sigma_k^\dagger\}$. We know the density operators $\{\sigma_k\,\widetilde{\rho}_i\,\sigma_k^\dagger\}$ are valid because σ_k is a unitary operator, and acting with a unitary operator such as σ_k implements a change of basis at the channel output. The question we are interested in is whether the output ensemble $\{p_i\,,\, \sigma_k\,\widetilde{\rho}_i\,\sigma_k^\dagger\}$ is *achievable*. To answer this, we know for each $\widetilde{\rho}_i$, there is a valid input ρ_i such that $\mathcal{E}(\rho_i) \;=\; \widetilde{\rho}_i$. Consider the following.

$$\sigma_k\,\widetilde{\rho}_i\,\sigma_k^\dagger \;=\; \sigma_k\,\mathcal{E}\,(\rho_i)\;\sigma_k^\dagger \;=\; \sigma_k\;\mathcal{E}\left(\frac{1}{2}\Big(I_2 \;+\; \omega_{x_i}\sigma_x \;+\; \omega_{y_i}\sigma_y \;+\; \omega_{z_i}\sigma_z\Big)\right)\;\sigma_k^\dagger$$

$$=\; \sigma_k\;\left(\frac{1}{2}\Big(I_2 \;+\; \lambda_x\omega_{x_i}\sigma_x \;+\; \lambda_y\omega_{y_i}\sigma_y \;+\; \lambda_z\omega_{z_i}\sigma_z\Big)\right)\;\sigma_k^\dagger$$

$$=\; \frac{1}{2}\Big(I_2 \;+\; \lambda_x\omega_{x_i}\sigma_k\sigma_x\sigma_k^\dagger \;+\; \lambda_y\omega_{y_i}\sigma_k\sigma_y\sigma_k^\dagger \;+\; \lambda_z\omega_{z_i}\sigma_k\sigma_z\sigma_k^\dagger\Big)\;.$$

Define $\bar{\delta}_{k,l} = 0$ if $k = l$, and 1 if $k \neq l$. Note that $\sigma_k \, \sigma_l \, \sigma_k = (-1)^{\bar{\delta}_{k,l}} \, \sigma_l$. If φ_i has the Bloch vector $\begin{bmatrix} (-1)^{\bar{\delta}_{k,x}} \, \omega_x \\ (-1)^{\bar{\delta}_{k,y}} \, \omega_y \\ (-1)^{\bar{\delta}_{k,z}} \, \omega_z \end{bmatrix}$, then the channel output of φ_i is

$$\mathcal{E}(\varphi_i) = \frac{1}{2} \Big(I_2 + (-1)^{\bar{\delta}_{k,x}} \, \lambda_x \omega_x \sigma_x + (-1)^{\bar{\delta}_{k,y}} \, \lambda_y \omega_y \sigma_y + (-1)^{\bar{\delta}_{k,z}} \, \lambda_z \omega_z \sigma_z \Big)$$

$$= \frac{1}{2} \Big(I_2 + \lambda_x \omega_x \sigma_k \sigma_x \sigma_k^\dagger + \lambda_y \omega_y \sigma_k \sigma_y \sigma_k^\dagger + \lambda_z \omega_z \sigma_k \sigma_z \sigma_k^\dagger \Big) = \sigma_k \, \mathcal{E} \, (\varphi_i) \, \sigma_k^\dagger. = \sigma_k \widetilde{\rho}_i \sigma_k^\dagger.$$

If we can show the φ_i are valid density operators, then we have shown that the output ensemble $\{p_i \, , \, \sigma_k \, \widetilde{\rho}_i \, \sigma_k^\dagger\}$ is achievable. In order for φ_i to be a valid density operator, we must have the corresponding Bloch vector composed of three real entries, and the magnitude of the Bloch vector less than or equal to one. Since the ρ_i are valid density operators, the three ω_k are real, and obey $\omega_x^2 + \omega_y^2 + \omega_z^2 \leq 1$. Now $(-1)^{\bar{\delta}_{k,l}} \; for \; k, l = \{x, y, z\}$ is real and equal in magnitude to one. The magnitude of the Bloch vector for φ_i is $\Big((-1)^{\bar{\delta}_{k,x}} \, \omega_x \Big)^2 + \Big((-1)^{\bar{\delta}_{k,y}} \, \omega_y \Big)^2 + \Big((-1)^{\bar{\delta}_{k,z}} \, \omega_z \Big)^2 = \omega_x^2 + \omega_y^2 + \omega_z^2 \leq 1$, where the last inequality follows from the fact that the ρ_i are valid density operators. Thus the φ_i are valid density operators. We conclude that if there exists an optimal input ensemble $\{p_i \, , \, \rho_i\}$, with corresponding output ensemble $\{p_i \, , \, \mathcal{E}(\rho_i) = \widetilde{\rho}_i\}$, then the ensemble $\left\{p_i \, , \, \sigma_k \widetilde{\rho}_i \sigma_k^\dagger\right\}$ is achievable, with corresponding input ensemble $\{p_i \, , \, \varphi_i\}$. Furthermore, the input ensemble $\{p_i \, , \, \varphi_i\}$ is optimal, since σ_k is a unitary operator, and we showed in equation II that a unitary operator acting on an ensemble does not change the Holevo quantity of that ensemble. Since $\{p_i \, , \, \mathcal{E}(\rho_i) = \widetilde{\rho}_i\}$attained the maximal Holevo quantity $\mathcal{C}_1$ at the channel output, the output ensemble $\left\{p_i \, , \, \sigma_k \widetilde{\rho}_i \sigma_k^\dagger\right\}$ also has a Holevo value of $\mathcal{C}_1$. Thus $\{p_i \, , \, \varphi_i\}$ is an optimal input ensemble.

To summarize, we first choose a basis of operators E_i, in this case the identity I_2 and the three Pauli operators $\{\sigma_x \, , \, \sigma_y \, , \, \sigma_z\}$, in which to expand the density matrix $\rho = \sum_i \alpha_i \, E_i$. Next, we found a set of unitary operators U_k, in this case again the Pauli operators σ_k, such that the U_k act on the E_i resulting in a multiplicative phase factor: $U_k E_i U_k^\dagger = \kappa_{(k,i)} \, E_i$, where $\kappa_{(k,i)}$ is a complex quantity. The unital nature of the qubit channel $\mathcal{E}$ tells us that $\mathcal{E}(E_i) = \lambda_i \, E_i \quad \forall \; i$ in the operator basis $\{E_i\}$. This leads to the commutation of the

channel $\mathcal{E}$ with the set of unitaries $\{U_k\} = \{ \pm I_2 , \pm\sigma_x , \pm\sigma_y , \pm\sigma_z\}$.

$$U_k \mathcal{E}(E_i) U_k^\dagger = U_k \lambda_i E_i U_k^\dagger = \lambda_i U_k E_i U_k^\dagger = \lambda_i \kappa_{(k,i)} E_i = \kappa_{(k,i)} \mathcal{E}(E_i)$$

$$(\textit{By linearity of quantum channels}) = \mathcal{E}\left(\kappa_{(k,i)} E_i\right) = \mathcal{E}\left(U_k E_i U_k^\dagger\right).$$

Since we have an expansion of ρ in terms of the E_i, using the linearity of quantum channels, we conclude that

$$U_k \mathcal{E}(\rho) U_k^\dagger = U_k \mathcal{E}\left(\frac{1}{2}\sum_i \alpha_i E_i\right) U_k^\dagger = U_k \left(\frac{1}{2}\sum_i \alpha_i \mathcal{E}(E_i)\right) U_k^\dagger \qquad \text{(III)}$$

$$= \frac{1}{2}\sum_i \alpha_i U_k \mathcal{E}(E_i) U_k^\dagger = \frac{1}{2}\sum_i \alpha_i \mathcal{E}\left(U_k E_i U_k^\dagger\right)$$

$$= \mathcal{E}\left(\frac{1}{2}\sum_i \alpha_i U_k E_i U_k^\dagger\right) = \mathcal{E}\left(U_k \left(\frac{1}{2}\sum_i \alpha_i E_i\right) U_k^\dagger\right) = \mathcal{E}\left(U_k \rho U_k^\dagger\right).$$

A U_k acting at the input is a basis change and hence $U_k \rho U_k^\dagger$ is a valid input density operator. Equation III allows us to conclude that any U_k acting on the output states $\widetilde{\rho}_i$ of an optimal ensemble $\{p_i , \rho_i\}$yields an output ensemble $\{p_i , U_k \widetilde{\rho}_i U_k^\dagger\}$ which is achievable. The achievability of channel output ensembles generated by U_k acting on the output ensemble of an optimal input ensemble will be a critical tool in extending the unital qubit channel analysis to the determination of the Holevo-Schumacher-Westmoreland channel capacity $\mathcal{C}_1$ for a special class of *qudit* unital channels.

IX.c Symmetry Properties of Optimal Ensembles

Consider a unital qubit channel with an optimal input ensemble $\{p_i, \rho_i\}$, [6] average input state $\Phi = \sum_i p_i \rho_i$ and average output state $\widetilde{\Phi} = \mathcal{E}(\Phi)$. Let Φ have Bloch vector $\vec{V} = \begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix}$ and $\widetilde{\Phi}$ have Bloch vector $\widetilde{\vec{V}} = \begin{bmatrix} \widetilde{v}_x \\ \widetilde{v}_y \\ \widetilde{v}_z \end{bmatrix} = \begin{bmatrix} \lambda_x v_x \\ \lambda_y v_y \\ \lambda_z v_z \end{bmatrix}$. Choose one of the three $\{\sigma_k\}$ and apply this σ_k to the output states $\widetilde{\rho}_i$ to obtain a new output

[6] That such an ensemble exists was shown in [16]. See property III in Section VI.a.

ensemble $\{p_i\,,\,\sigma_k\widetilde{\rho}_i\sigma_k^\dagger\}\;\equiv\;\{p_i\,,\,\widetilde{\rho}_i'\}$. We know from the work above that the output ensemble $\{p_i\,,\,\sigma_k\widetilde{\rho}_i\sigma_k^\dagger\}$ is achievable and optimal. The action of σ_k on the output ensemble $\{p_i\,,\,\mathcal{E}(\rho_i)\;=\;\widetilde{\rho}_i\}$generates a corresponding transformation of the average output state of the optimal ensemble $\widetilde{\Phi}$,

$$\sum_i p_i\,\sigma_k\,\widetilde{\rho}_i\,\sigma_k^\dagger\;=\;\sigma_k\left(\sum_i p_i\,\widetilde{\rho}_i\right)\sigma_k^\dagger\;=\;\sigma_k\widetilde{\Phi}\sigma_k^\dagger\;=\;\widetilde{\Phi}'.$$

By the invariance property shown in Section VIII.b, we have $\widetilde{\Phi}'\;\equiv\;\widetilde{\Phi}$. Now $\widetilde{\Phi}$ has the Bloch vector $\widetilde{\vec{V}}\;=\;\begin{bmatrix}\widetilde{v}_x\\ \widetilde{v}_y\\ \widetilde{v}_z\end{bmatrix}\;=\;\begin{bmatrix}\lambda_x\,v_x\\ \lambda_y\,v_y\\ \lambda_z\,v_z\end{bmatrix}$ and $\widetilde{\Phi}'$ has the Bloch vector $\widetilde{\vec{V}}'\;=\;\begin{bmatrix}(-1)^{\bar{\delta}_{k,x}}\;\widetilde{v}_x\\ (-1)^{\bar{\delta}_{k,y}}\;\widetilde{v}_y\\ (-1)^{\bar{\delta}_{k,z}}\;\widetilde{v}_z\end{bmatrix}$. For $k=\{x,y,z\}\;,\;\widetilde{\Phi}\;\equiv\;\widetilde{\Phi}'$ implies

$$\widetilde{v}_x\;=\;(-1)^{\bar{\delta}_{k,x}}\;\widetilde{v}_x\qquad and\qquad \widetilde{v}_y\;=\;(-1)^{\bar{\delta}_{k,y}}\;\widetilde{v}_y\qquad and\qquad \widetilde{v}_z\;=\;(-1)^{\bar{\delta}_{k,z}}\;\widetilde{v}_z. \qquad \text{(IV)}$$

The only way all three relationships in equation IV can be true $\forall\;k=\{x,y,z\}$ is if $\widetilde{v}_x=\widetilde{v}_y=\widetilde{v}_z=0$. The fact that $\widetilde{\Phi}$ has the Bloch vector $\widetilde{\vec{V}}\;=\;\begin{bmatrix}0\\0\\0\end{bmatrix}$ leads to the conclusion that $\widetilde{\Phi}\;=\;\frac{1}{2}\left(I_2\;+\;\widetilde{\vec{V}}\bullet\vec{\sigma}\right)\;=\;\frac{1}{2}\,I_2$ for all optimal ensembles.

A second way to see that $\widetilde{\Phi}\;\equiv\;\frac{1}{2}\,I_2$ is via Schur's Lemma[27]. Consider the group $\mathcal{H}$ composed of the eight operations $\{\,\pm I_2\,,\,\pm\sigma_x\,,\,\pm\sigma_y\,,\,\pm\sigma_z\,\}$, and a two dimensional representation $\Gamma(\mathcal{H})$ of $\mathcal{H}$. A necessary and sufficient condition for a representation $\Gamma(\mathcal{H})$ of a finite group to be irreducible is if the relation $\frac{1}{\|\mathcal{H}\|}\sum_{h\in\mathcal{H}}\left|Trace[\Gamma(h)]\right|^2\;=\;1$ is true[27]. Here $\|\mathcal{H}\|$ is the order of the group $\mathcal{H}$. The group $\mathcal{H}$ is finite, with $\|\mathcal{H}\|=8$. Computing the trace sum with the standard two dimensional Pauli matrices for the representation of $\mathcal{H}$, we find the qubit Pauli based representation of $\mathcal{H}$ is irreducible.

Schur's Lemma states that if a group $\mathcal{H}$ has a d-dimensional irreducible representation $\Gamma(\mathcal{H})$ such that each representation element $\Gamma(h)$ commutes with a d by d matrix M, $\forall\,h\;\in\mathcal{H}$, then M is proportional to the d by d identity matrix I_d[27]. The fact that $\sigma_k\widetilde{\Phi}\sigma_k^\dagger\;=$

$\widetilde{\Phi} \quad \forall k \in \{x, y, z\}$, together with the same trivial result for I_2, implies that all the qubit representation elements $\Gamma(h)$ of $\mathcal{H}$ commute with $\widetilde{\Phi}$ and thus $\widetilde{\Phi} \propto I_2$. The trace condition $Trace\left(\widetilde{\Phi}\right) = 1$ leads us to conclude $\widetilde{\Phi} = \frac{1}{2} I_2$.

Having determined $\widetilde{\Phi} = \frac{1}{2} I_2$, note that $\mathcal{S}\left(\widetilde{\Phi}\right) = \log_2(2) = 1$. Using this result, we rewrite the Holevo-Schumacher-Westmoreland channel capacity as $\mathcal{C}_1 = 1 - \sum_i p_i \mathcal{S}\left(\mathcal{E}\left(\rho_i\right)\right)$. To further simplify $\mathcal{C}_1$, we use two properties from Section VI.a, rewritten in the notation of this section.

I) *The equal distance property of optimal ensembles.*
For any optimal ensemble $\{p_i\,,\,\rho_i\}$, we have

$$\mathcal{D}\Big[\mathcal{E}(\rho_i)\Big\|\mathcal{E}(\Phi)\Big] = \mathcal{C}_1 \quad \forall i. \tag{V}$$

II) *The sufficiency of the maximal distance property.*
For any optimal ensemble $\{p_i\,,\,\rho_i\}$with average input state $\Phi = \sum_i p_i \rho_i$, we have

$$\mathcal{D}\Big[\mathcal{E}(\phi)\Big\|\mathcal{E}(\Phi)\Big] \leq \mathcal{C}_1 \quad for \quad any \quad input \quad density \quad matrix \quad \phi. \tag{VI}$$

In both I) and II), $\Phi = \sum_i p_i \rho_i$. For the case of qubit unital channels, we have found that every optimal ensemble $\{p_i\,,\,\rho_i\}$must obey $\mathcal{E}\Big(\sum_i p_i\,\rho_i\Big) = \frac{1}{2} I_2$. Looking at the relative entropy formula, we see that $\mathcal{D}\Big[\mathcal{E}\left(\phi\right)\Big\|\frac{1}{d}\,I_d\Big] = \log_2(d) - \mathcal{S}\left(\mathcal{E}\left(\phi\right)\right)$, and ϕ is any input density matrix. Using the fact that for qubit unital channels we have found, for all optimal ensembles $\{p_i\,,\,\rho_i\}$, that $\mathcal{E}\Big(\sum_i p_i\,\rho_i\Big) = \frac{1}{2} I_2$, the above two Schumacher and Westmoreland results become, in the qubit unital channel case,

I')

$$1 - \mathcal{S}\left(\mathcal{E}(\rho_i)\right) = \mathcal{C}_1 \quad \forall i \qquad implying \quad \mathcal{S}\left(\mathcal{E}\left(\rho_i\right)\right) = \mathcal{S}\left(\mathcal{E}\left(\rho_j\right)\right)\ \forall\, i, j. \tag{VII}$$

II')

$$1 - \mathcal{S}\left(\mathcal{E}(\phi)\right) \leq \mathcal{C}_1 \quad \forall \quad input \quad density \quad matrices \quad \phi. \tag{VIII}$$

We know that II') is achieved with equality when ϕ is any of the ρ_i in the optimal ensemble $\{p_i\,,\,\rho_i\}$. Thus I') and II') taken together yield $1 - \mathcal{S}\left(\mathcal{E}(\phi)\right) \leq 1 - \mathcal{S}\left(\mathcal{E}(\rho_i)\right)$ or $\mathcal{S}\left(\mathcal{E}(\phi)\right) \geq \mathcal{S}\left(\mathcal{E}(\rho_i)\right)$, which, since ϕ can be any input density matrix, implies $\mathcal{S}\left(\mathcal{E}(\rho_i)\right) = min_\phi\, \mathcal{S}\left(\mathcal{E}(\phi)\right)$. Plugging this result into I') yields our final result for the Holevo-Schumacher-Westmoreland channel capacity for qubit unital channels:

$$\mathcal{C}_1 \;=\; 1 \,-\, min_\phi\, \mathcal{S}\left(\mathcal{E}(\phi)\right).$$

For qubit unital channels, the minimum channel output von Neumann entropy determines the Holevo-Schumacher-Westmoreland channel capacity $\mathcal{C}_1$.

IX.d Ensemble Achievability

The achievability of a transformed output ensemble is a concept worth emphasizing. In our discussion of unital qubit channels, the reason why we could conclude the average output state of an optimal ensemble commuted with all eight members of the representation Γ of the group $\mathcal{H} = \{\pm I_2\,, \pm\sigma_x\,, \pm\sigma_y\,, \pm\sigma_z\}$ was because, given an optimal ensemble $\{p_i\,,\,\rho_i\}$, each of the eight output ensembles $\{p_i\,,\,\Gamma(h)\widetilde{\rho}_i\Gamma(h^{-1})\}$, where $h \in \mathcal{H}$, was achievable. The existence of an optimal input ensemble $\{p_i\,,\,\phi_i\}$which maps via the quantum channel $\mathcal{E}$ to $\{p_i\,,\,\Gamma(h)\widetilde{\rho}_i\Gamma(h^{-1})\}$ is what allowed us to conclude the relationship $\Gamma(h)\widetilde{\Phi}\Gamma(h^{-1}) = \widetilde{\Phi}$ was valid $\forall\, h \in \mathcal{H}$, and apply Schur's Lemma.

For a generic group $\mathcal{M}$ and corresponding irreducible representation $\Pi(\mathcal{M})$ acting on the channel output of an optimal ensemble $\{p_i\,,\,\rho_i\}$, there will typically be $m_0 \in \mathcal{M}$ such that $\left\{p_i\,,\,\Pi(m_0)\widetilde{\rho}_i\Pi(m_0^{-1})\right\}$ are not achievable ensembles. In these cases, we cannot conclude $\Pi(m_0)\widetilde{\Phi}\Pi(m_0^{-1}) = \widetilde{\Phi}$ holds, where $\widetilde{\Phi}$ is the average output state of an optimal ensemble. Yet it was the fact that $\Pi(m_0)\widetilde{\Phi}\Pi(m_0^{-1}) = \widetilde{\Phi}$ holds $\forall\, m \in \mathcal{M}$ that led us to apply Schur's

Lemma and conclude $\widetilde{\Phi} \propto I_d$. The lack of achievability for one or more of the transformed output ensembles $\left\{p_i, \Pi(m)\,\widetilde{\rho}_i\,\Pi(m^{-1})\right\}$ prevents us from appealing to Schur's Lemma. An example of the limitations to determining the HSW channel capacity which results from output ensemble non-achievability arises in the case of non-unital qubit channels.

IX.e A Non-Unital Qubit Channel Example

Our technique fails for non-unital qubit channels. The reason why is the lack of achievability of output ensembles generated by members of the Pauli group acting on an output optimal ensemble. For example, consider the non-unital linear qubit channel specified in the Ruskai-King-Swarez-Werner notation as $\{t_x = t_y = 0, t_z = 0.2, \lambda_x = \lambda_y = 0, \lambda_z = 0.4\}$. This channel maps an input Bloch vector $\vec{W}$ to an output Bloch vector $\widetilde{\vec{W}}$ as:

$$\vec{W} = \begin{bmatrix} w_x \\ w_y \\ w_z \end{bmatrix} \longrightarrow \begin{bmatrix} 0 \\ 0 \\ t_z + \lambda_z w_z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0.2 + 0.4 w_z \end{bmatrix} = \widetilde{\vec{W}}.$$

By inspection, an optimal input ensemble is $\{p_i\,,\,\rho_i\}$with $\rho_{1,2} = \frac{1}{2}\,(I_2 \pm \sigma_z)$, and corresponding output density matrices $\widetilde{\rho_1} = \frac{1}{2}\,(I_2 - 0.2\sigma_z)$ and $\widetilde{\rho_2} = \frac{1}{2}\,(I_2 + 0.6\sigma_z)$. Numerical analysis for this channel indicates the optimum output average state is $\widetilde{\Phi} \approx \frac{1}{2}\,(I_2 + 0.2125\,\sigma_z)$. Since $\widetilde{\Phi} \neq \frac{1}{2}\,I_2$, we anticipate we will not be able to meet the conditions for the application of Schur's Lemma.

Consider applying the unitary operator σ_z to the output optimal ensemble $\{p_i\,,\,\mathcal{E}(\rho_i) = \widetilde{\rho}_i\}$determined in the previous paragraph. We obtain

$$\sigma_z\,\widetilde{\rho_1}\,\sigma_z = \sigma_z\left(\frac{1}{2}\left(I_2 - 0.2\sigma_z\right)\right)\sigma_z^\dagger = \widetilde{\rho_1}$$

and

$$\sigma_z\,\widetilde{\rho_2}\,\sigma_z = \sigma_z\left(\frac{1}{2}\left(I_2 + 0.6\sigma_z\right)\right)\sigma_z^\dagger = \widetilde{\rho_2}.$$

Thus the output ensemble $\left\{p_i\,,\,\sigma_z\,\mathcal{E}\,(\rho_i)\;\sigma_z^\dagger = \sigma_z\,\widetilde{\rho}_i\,\sigma_z^\dagger\right\}$ is identical to the output ensemble

$\left\{p_i\,,\,\mathcal{E}\left(\rho_i\right)\,=\,\widetilde{\rho}_i\right\}$, both being generated by the input ensemble $\{p_i\,,\,\rho_i\}$. Thus the output ensemble $\left\{p_i\,,\,\sigma_z\,\mathcal{E}\left(\rho_i\right)\,\sigma_z^\dagger\,=\,\sigma_z\,\widetilde{\rho}_i\,\sigma_z^\dagger\right\}$ is an achievable output ensemble.

The application of σ_x or σ_y to $\{p_i\,,\,\mathcal{E}(\rho_i)\,=\,\widetilde{\rho}_i\}$however does not yield an achievable ensemble. To see why, consider applying σ_x to $\widetilde{\rho_2}\,=\,\frac{1}{2}\left(I_2\,+\,0.6\sigma_z\right)$, which since $\sigma_x\sigma_z\sigma_x^\dagger\,=\,-\sigma_z$, yields the output density operator $\widetilde{\rho_2}'\,=\,\frac{1}{2}\left(I_2\,-\,0.6\sigma_z\right)$. The corresponding input density operator would have Bloch vector $\vec{W}'\,=\,\begin{bmatrix}0\\0\\-2\end{bmatrix}$, which is not a valid qubit density operator, since $\|\vec{W}'\|\,>\,1$. Since the output state $\sigma_x\widetilde{\rho_2}\sigma_x^\dagger$ can never be mapped to by a valid input qubit density operator, we cannot assume the relation $\sigma_x\,\widetilde{\Phi}\,\sigma_x^\dagger\,=\,\widetilde{\Phi}$ holds. A similar analysis for σ_y indicates we also cannot assume the relation $\sigma_y\,\widetilde{\Phi}\,\sigma_y^\dagger\,=\,\widetilde{\Phi}$ holds.

Thus, we do not have the necessary Schur commutation requirement that $\Gamma(g)\widetilde{\Phi}\,=\,\widetilde{\Phi}\Gamma(g)$ for all representation elements Γ of the group $\mathcal{G}=\{\pm I_2,\pm\sigma_x,\pm\sigma_y,\pm\sigma_z\}$, and hence cannot conclude $\widetilde{\Phi}\,=\,\frac{1}{2}I_2$, as we anticipated. As we shall develop in more detail below, working with qudits, if we can find a group $\mathcal{G}$ with a d-dimensional representation Γ which is unitary and irreducible, such that $\Gamma(g)$ acting on the output states of an optimal ensemble $\{p_i\,,\,\rho_i\}$yield achievable ensembles $\forall\,g\,\in\,\mathcal{G}$, than we will be able to conclude the average output state of any optimal ensemble is $\widetilde{\Phi}\,=\,\frac{1}{d}\,I_d$. From this conclusion, we can use the Schumacher-Westmoreland relative entropy properties from Section VI.a, as embodied in equations V,VI,VII, and VIII, to conclude the states in any input optimal ensemble must be a subset of those input states which yield the minimum output von Neumann entropy. This in turn leads to a HSW channel capacity of

$$\mathcal{C}_1\;=\;\log_2(d)\;-\;min_\phi\,\mathcal{S}\left(\mathcal{E}(\phi)\right)$$

for those qudit channels to which we can successfully apply Schur's Lemma. We now proceed to determine the subset of qudit channels which meet the Schur's Lemma requirements.

X Qudit Channels

The HSW channel capacity result for unital qubit channels was previously proven in [25] by a method which did not generalize to the general qudit case (i.e.: for qudit dimension $d > 2$). The technique discussed in this chapter does generalize to a special subclass of unital qudit channels. Before describing that generalization, we present some background material on qudits and qudit channels.

X.a Qudits

A qudit is a system with d orthogonal pure states $|j\rangle$, $j = 0, 1, 2, \cdots, d-1$. The generalization of the qubit Pauli operators σ_x and σ_z are the two operators $\hat{X}$ and $\hat{Z}$, whose action on the states $|j\rangle$ are $\hat{X}|j\rangle = |j+1\,(mod\ d)\rangle$ and $\hat{Z}|j\rangle = \Omega^j\,|j\rangle$. Here $\Omega = e^{\frac{2\pi i}{d}}$. The extension of the qubit Bloch representation for a density matrix ρ to qudits is shown in appendix B to be

$$\rho = \frac{1}{d} \sum_{a,b \in \{0,1,2,\cdots,d-1\}} \alpha_{a,b}\ \hat{X}^a\ \hat{Z}^b.$$

The $\alpha_{a,b}$ are complex quantities. Define $E_{a,b} = \hat{X}^a \hat{Z}^b$. Note that $E_{0,0} = I_d$. In appendix B it is shown $Trace(E_{a,b}) = d\,\delta_{a,0}\,\delta_{b,0}$, where δ is the Kronecker delta function. The trace condition $Trace(\rho) = 1$ allows us to conclude $\alpha_{0,0} = 1$. Let Υ denote the set of $d^2 - 1$ elements $a, b \in \{0, 1, 2, \cdots, d-1\}$ *with the exception that a and b cannot both be zero.* Then we can write the qudit density matrix ρ as $\rho = \frac{1}{d}\left(I_d + \sum_{(a,b)\in\Upsilon}\ \alpha_{a,b}\ E_{a,b}\right)$. A qudit quantum channel $\mathcal{E}$ is a linear map. One can write such a map as a d^2 by d^2 complex matrix $\mathcal{M}$ taking the d^2 vector of coefficients $\alpha_{a,b}$ of ρ to the d^2 set of coefficients $\widetilde{\alpha}_{a,b}$ of $\widetilde{\rho} = \mathcal{E}(\rho)$. [7]

If the qudit quantum channel $\mathcal{E}$ is unital, meaning $\mathcal{E}(I_d) = I_d$, then the first row and column of $\mathcal{M}$ must be a one followed by $d^2 - 1$ zeros. Hence we can represent a qudit unital

[7] Our qudit matrix development in which we write $\mathcal{E}$ as a d^2 by d^2 matrix closely follows work done in [26] for the unital qubit channel case.

channel by a matrix $\mathcal{N}$ of $d^2 - 1$ by $d^2 - 1$ complex entries mapping the vector of $d^2 - 1$ coefficients $\alpha_{(a,b)}$, with $(a,b) \in \Upsilon$, representing ρ to the vector of $d^2 - 1$ coefficients $\widetilde{\alpha}_{(a,b)}$, with $(a,b) \in \Upsilon$, representing $\widetilde{\rho} = \mathcal{E}(\rho)$. The specific class of qudit channels we shall be interested in are those completely positive unital quantum channels for which $\mathcal{N}$ is diagonal. This class of channels is nonempty. For example, consider the channel corresponding to all zeros on the diagonal. This point channel maps all input density matrices to a single output density matrix $\widetilde{\rho} = \frac{1}{d} I_d$. Another member of the set of diagonal unital channels is the identity map, which maps any input density matrix to itself. This channel has all ones on the diagonal of the matrix $\mathcal{N}$.

X.b The King-Ruskai-Szarek-Werner Qubit Channel Representation Proof and Higher Dimensional Systems

The King-Ruskai-Szarek-Werner qubit channel representation proof does not extend to systems with dimension $d > 2$. To see why, recall the King-Ruskai-Szarek-Werner derivation, briefly outlined below. (See [25] for more detail.)

Consider qubits in the Bloch vector representation. Let $\rho = \frac{1}{2} (I_2 + \vec{\mathcal{W}}_\rho \bullet \vec{\sigma})$. For a general, possibly non-unital, qubit channel, the map $\mathcal{E}$ acting on ρ can be represented in terms of a map of the Bloch vector, as shown below.

$$\mathcal{E}(\rho) = \frac{1}{2} \left(I_2 + \left(\mathbf{T} \vec{\mathcal{W}}_\rho + \vec{t} \right) \bullet \vec{\sigma} \right), \tag{IX}$$

where we use the notation of [25]. Here $\mathbf{T}$ is a real three by three matrix, and $\vec{t}$ is a real three element vector. Using the polar decomposition, we can write $\mathbf{T}$ as a rotation matrix $\mathbf{R}$ times a self - adjoint matrix $\mathbf{S}$, yielding $\mathbf{T} = \mathbf{R}\,\mathbf{S}$. Equation IX becomes

$$\mathcal{E}(\rho) = \frac{1}{2} \left(I_2 + \left(\mathbf{R}\mathbf{S} \vec{\mathcal{W}}_\rho + \vec{t} \right) \bullet \vec{\sigma} \right) = \frac{1}{2} \left(I_2 + \mathbf{R} \left(\mathbf{S} \vec{\mathcal{W}}_\rho + \mathbf{R}^{-1} \vec{t} \right) \bullet \vec{\sigma} \right).$$

Since $\mathbf{S}$ is self - adjoint, it can be diagonalized, yielding in a new basis the real diagonal

matrix $\mathbf{D}$. In this new Bloch vector basis, the primed basis, we have:

$$\mathcal{E}(\rho) \;=\; \frac{1}{2}\left(I_2 \;+\; \left[\acute{\mathbf{R}} \left(\mathbf{D}\, \acute{\vec{\mathcal{W}}}_\rho \;+\; \acute{\mathbf{R}}^{-1}\, \acute{\vec{t}} \right) \right] \;\bullet\; \vec{\sigma} \right).$$

Recall that the group space of $SU(2)$ is diffeomorphic to the three sphere, S^3[28]. This allows us to replace the rotation $\acute{\mathbf{R}}$ acting on the Bloch vector with a special unitary U acting on the density matrix ρ. Denoting the vector of real elements $\acute{\mathbf{R}}^{-1}\, \acute{\vec{t}}$ as $\widetilde{\vec{t}}$, we have:

$$\mathcal{E}(\rho) \;=\; U \left[\frac{1}{2}\left(I_2 \;+\; \left(\mathbf{D}\, \acute{\vec{\mathcal{W}}}_\rho \;+\; \widetilde{\vec{t}} \right) \bullet\; \vec{\sigma} \right) \right] U^\dagger.$$

Thus, up to a special unitary acting on the output of the channel, there exists a basis for the Bloch vector of the input density matrix, such that the channel can be thought of as a rescaling along the σ_x, σ_y, and σ_z operator basis axes by the diagonal elements of $\mathbf{D}$, and a shifting of these same axes by the elements of the vector $\widetilde{\vec{t}}$. Thus the action of the channel on the Bloch vector $\vec{\mathcal{W}}$ can be considered to be:

$$\mathcal{W}_k \;\longrightarrow\; t_k \;+\; \lambda_k\, \mathcal{W}_k \qquad for \quad k \;=\; x, y, z\ .$$

For the general, d - dimensional qudit case, the derivation for qubits fails at the step where the rotation $\mathbf{R}$ is replaced by a special unitary U. Hence we cannot, at least by this method, think of general qudit channels as a rescaling and a shift of the Generalized Pauli operator basis elements.

In the work that follows, we will consider general qudit channels in the Generalized Pauli operator basis. However, because the King-Ruskai-Szarek-Werner qubit channel representation does not extend to qudit channels, we shall develop a different, algebraicly motivated analysis for diagonal, unital qudit channels.

X.c Extending the Qubit Unital Channel Analysis to Diagonal Unital Qudit Channels

The approach we take to determine the HSW channel capacity for the class of diagonal unital qudit channels closely follows our unital qubit channel derivation above. Note the operators $E_{a,b}$ are unitary. Using the commutation relation shown in appendix B, $\hat{Z}\hat{X} = \Omega\hat{Z}\hat{X}$, where $\Omega = e^{\frac{2\pi i}{d}}$, we have

$$E_{g,h}\, E_{a,b}\, E_{g,h}^{\dagger} = \hat{X}^g \hat{Z}^h\, \hat{X}^a \hat{Z}^b\, \hat{Z}^{-h} \hat{X}^{-g} = \Omega^{ah}\, \hat{X}^g \hat{X}^a\, \hat{Z}^h \hat{Z}^b\, \hat{Z}^{-h} \hat{X}^{-g} \tag{X}$$

$$= \Omega^{ah}\, \hat{X}^g \hat{X}^a \hat{Z}^b \hat{X}^{-g} = \Omega^{ah}\, \Omega^{-bg}\, \hat{X}^g \hat{X}^a \hat{X}^{-g} \hat{Z}^b = \Omega^{ah}\, \Omega^{-bg}\, \hat{X}^a \hat{X}^g \hat{X}^{-g} \hat{Z}^b$$

$$= \Omega^{ah}\, \Omega^{-bg}\, \hat{X}^a \hat{Z}^b = \Omega^{ah\,-\,bg}\, E_{a,b}.$$

Define $F_{a,b,c} = \Omega^c\, E_{a,b}$, where $a, b, c \in \{0, 1, 2, \cdots, d-1\}$. Since Ω^c and the $E_{a,b}$ are unitary operators, $F_{a,b,c}$ is a unitary operator. The action of the $F_{a,b,c}$ on a diagonal unital qudit channel output density operator $\widetilde{\rho}$ is

$$F_{a,b,c}\, \widetilde{\rho}\, F_{a,b,c}^{\dagger} = E_{a,b}\, \widetilde{\rho}\, E_{a,b}^{\dagger} = E_{a,b}\, \mathcal{E}(\rho)\, E_{a,b}^{\dagger} \tag{XI}$$

$$= E_{a,b} \frac{1}{d} \left(I_d + \sum_{(q,r)\in\Upsilon} \lambda_{q,r} \alpha_{q,r} E_{q,r} \right) E_{a,b}^{\dagger} = \frac{1}{d} \left(I_d + \sum_{(q,r)\in\Upsilon} \lambda_{q,r} \alpha_{q,r} E_{a,b} E_{q,r} E_{a,b}^{\dagger} \right)$$

$$= \frac{1}{d} \left(I_d + \sum_{(q,r)\in\Upsilon} \lambda_{q,r}\, \alpha_{q,r}\, \Omega^{bq-ar}\, E_{q,r} \right) = \mathcal{E}\left(\frac{1}{d} \left(I_d + \sum_{(q,r)\in\Upsilon} \alpha_{q,r}\, \Omega^{bq-ar}\, E_{q,r} \right) \right)$$

$$= \mathcal{E}\left(E_{a,b} \frac{1}{d} \left(I_d + \sum_{(q,r)\in\Upsilon} \alpha_{q,r}\, E_{q,r} \right) E_{a,b}^{\dagger} \right) = \mathcal{E}\left(E_{a,b}\, \rho\, E_{a,b}^{\dagger} \right) = \mathcal{E}\left(F_{a,b}\, \rho\, F_{a,b,c}^{\dagger} \right).$$

Since the $F_{a,b,c}$ are unitary operators, we conclude that given any optimal input ensemble $\{p_i\,,\, \rho_i\}$, the output ensemble $\Theta_{a,b,c}$ obtained by applying $F_{a,b,c}$ to $\{p_i\,,\, \mathcal{E}(\rho_i) = \widetilde{\rho}_i\}$is achievable and $\Theta_{a,b,c}$ has the optimal input ensemble $\left\{p_i\,,\, \phi_i = F_{a,b,c}\, \rho_i\, F_{a,b,c}^{\dagger}\right\}$. Each of the ϕ_i is a valid input density operator due to the fact that $F_{a,b,c}$ is a unitary operator and is implementing a change of basis on ρ_i.

The set of operators $\{F_{a,b,c}\}$ form a finite group of order d^3 which we shall call $\mathcal{Q}$. Consider a d-dimensional representation Γ of $\mathcal{Q}$ constructed using d-dimensional matrix representations for the operators $\hat{X}$ and $\hat{Z}$ acting on the d kets $|j\rangle$, $j = 0, \cdots, d-1$. [8] Recall the theorem for proving a representation Γ of a finite group is irreducible[27]. The group representation Γ of $\mathcal{Q}$ is irreducible since $\left| Trace\left[F_{a,b,c}\right] \right|$ equals zero when either a and b are non-zero, and $\left| Trace\left[F_{a,b,c}\right] \right|$ equals d when $a = b = 0$. Thus $\frac{1}{\|\mathcal{Q}\|} \sum_{q \in \mathcal{Q}} \left| Trace\left[\Gamma(q)\right] \right|^2 = \frac{1}{d^3}\, d\, d^2 = 1$. Since $\Gamma(\mathcal{Q})$ is irreducible, we can apply Schur's Lemma. For any optimal input ensemble $\{p_i\,,\ \rho_i\}$, the channel output ensemble $\left\{p_i\,,\ F_{a,b,c}\,\mathcal{E}(\rho_i)\,F_{a,b,c}^\dagger\right\}$ is achievable and the corresponding input ensemble $\left\{p_i\,,\ F_{a,b,c}\,\rho_i\,F_{a,b,c}^\dagger\right\}$ is optimal. From the uniqueness of the average output state $\widetilde{\Phi}$ for any optimal ensemble, we conclude that $\forall\ a, b, c\ :\ F_{a,b,c}\,\widetilde{\Phi}\,F_{a,b,c}^\dagger = \widetilde{\Phi}$ or $F_{a,b,c}\,\widetilde{\Phi} = \widetilde{\Phi}\,F_{a,b,c}$. By Schur's Lemma we obtain $\widetilde{\Phi} \propto I_d$. The trace condition tells us $Trace\left(\widetilde{\Phi}\right) = 1$, so we conclude $\widetilde{\Phi} \equiv \frac{1}{d}\, I_d$.

This leads us to conclude that for the optimal input ensemble $\{p_i\,,\ \rho_i\}$, the HSW channel capacity is $\mathcal{C}_1 = \log_2(d) - \sum_i p_i \mathcal{S}(\mathcal{E}(\rho_i))$. Using the relative entropy properties in Section VI.a, as embodied in equations V,VI,VII, and VIII, we obtain $\mathcal{S}(\mathcal{E}(\rho_i)) = min_\phi\, \mathcal{S}\left(\mathcal{E}(\phi)\right)$ yielding the HSW channel capacity for diagonal unital qudit channels:

$$\mathcal{C}_1 = \log_2(d) - min_\phi\, \mathcal{S}\left(\mathcal{E}(\phi)\right).$$

XI Products of Diagonal Unital Qudit Channels

Consider the product of N diagonal unital qudit channels $\mathcal{E}^{(k)}$, $k = 1, \cdots, N$. The tensor product channel is $\mathcal{E}^{\otimes} = \mathcal{E}^{(1)} \otimes \mathcal{E}^{(2)} \otimes \cdots \otimes \mathcal{E}^{(N)}$. Let the input qudit density operator $\rho^{(k)}$ corresponding to the diagonal unital channel $\mathcal{E}^{(k)}$ be of dimension d_k. Then $d = \prod_{k=1}^N d_k$ is the dimension of the input qudit density matrix $\rho^{\otimes}$ for the product channel $\mathcal{E}^{\otimes}$. The basis elements for $\rho^{\otimes}$ which we shall use are the tensor products of the individual $E_{a,b}^{(k)}$.

$$\left\{E_{a,b}^{\otimes}\right\} = \left\{E_{a_1,b_1}^{(1)} \otimes E_{a_2,b_2}^{(2)} \otimes \cdots \otimes E_{a_N,b_N}^{(N)}\right\},$$

[8] See Section XII for $d = 3$ examples of the matrices for $\hat{X}$ and $\hat{Z}$.

a_N	0	0	0	$\cdots$	0	0	0	0	$\cdots$	0	0	0	$\cdots$
a_{N-1}	0	0	0	$\cdots$	0	0	0	0	$\cdots$	0	0	0	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$
a_3	0	0	0	$\cdots$	0	0	0	0	$\cdots$	0	0	0	$\cdots$
a_2	0	0	0	$\cdots$	0	1	1	1	$\cdots$	1	2	2	$\cdots$
a_1	0	1	2	$\cdots$	d_1-1	0	1	2	$\cdots$	d_1-1	0	1	$\cdots$
a	0	1	2	$\cdots$	d_1-1	d_1	d_1+1	d_1+2	$\cdots$	$2d_1-1$	$2d_1$	$2d_1+1$	$\cdots$

Table 1.2: Table of indices for the tensor product basis.

where the a_k and $b_k \in \{0, 1, 2, \ldots, d_k - 1\}$ and $(a, b) \in \{0, 1, 2, \ldots, d - 1\}$.

The basis elements $E^{\otimes}_{a,b}$ are not necessarily constructed using the d dimensional qudit operators $\hat{X}$ and $\hat{Z}$ described in appendix B. As a result, we must prove several properties for the basis set $\left\{E^{\otimes}_{a,b}\right\}$ before we proceed with the HSW channel capacity analysis for product channels.

XI.a The Relation between the Basis $\{E^{(1)}_{a_1,b_1} \otimes E^{(2)}_{a_2,b_2} \otimes \cdots \otimes E^{(N)}_{a_N,b_N}\}$ and the Basis $\{E^{\otimes}_{a,b}\}$

The set of basis elements $\{E^{(1)}_{a_1,b_1} \otimes E^{(2)}_{a_2,b_2} \otimes \cdots \otimes E^{(N)}_{a_N,b_N}\}$ and the set of basis elements $\{E^{\otimes}_{a,b}\}$ both have d elements, where $d = \prod_{k=1}^{k=N} d_k$. Here the $\{a^{(k)}_i, b^{(k)}_i\} \in \{0, 1, 2, \cdots, d_k - 1\}$ and the $\{a^{\otimes}, b^{\otimes}\} \in \{0, 1, 2, \cdots, d\}$. There are many bijective mappings between these two sets, and it is useful to have one particular map in mind as we proceed. The one we shall use is presented in the table below.

Below we associate an $E^{\otimes}_{a,b}$ with the tensor product $\left\{E^{(1)}_{a_1,b_1} \otimes E^{(2)}_{a_2,b_2} \otimes \cdots \otimes E^{(N)}_{a_N,b_N}\right\}$ by using this mapping twice, once for the association $\left\{a^{(1)}_{i_1}, a^{(2)}_{i_2}, a^{(3)}_{i_3}, \cdots, a^{(N-1)}_{i_{N-1}}, a^{(N)}_{i_N}\right\} \Longleftrightarrow \{a^{\otimes}\}$ and again for $\left\{b^{(1)}_{i_1}, b^{(2)}_{i_2}, b^{(3)}_{i_3}, \cdots, b^{(N-1)}_{i_{N-1}}, b^{(N)}_{i_N}\right\} \Longleftrightarrow \{b^{\otimes}\}$.

XI.b Orthonormality of the $\{E^{\otimes}_{a,b}\}$

The operators $E^{\otimes}_{a,b}$ form, with respect to the Hilbert-Schmidt norm, a set of d^2 orthogonal operators. The orthogonality of the $\left\{E^{\otimes}_{a,b}\right\}$ is inherited from the orthogonality of the operators $\left\{E^{(k)}_{a_k,b_k}\right\}$, which is shown in appendix B, equation XVII. Using properties of tensors from [29], we have

$$\left\langle E^{\otimes}_{a,b}, E^{\otimes}_{g,h} \right\rangle = Trace\left[E^{\otimes\dagger}_{a,b}\, E^{\otimes}_{g,h}\right] \tag{XII}$$

$$= Trace\left[\left(E^{(1)}_{a_1,b_1} \otimes E^{(2)}_{a_2,b_2} \otimes \cdots \otimes E^{(N)}_{a_N,b_N}\right)^{\dagger} \left(E^{(1)}_{g_1,h_1} \otimes E^{(2)}_{g_2,h_2} \otimes \cdots \otimes E^{(N)}_{g_N,h_N}\right)\right]$$

$$= Trace\left[\left(E^{(1)\dagger}_{a_1,b_1} \otimes E^{(2)\dagger}_{a_2,b_2} \otimes \cdots \otimes E^{(N)\dagger}_{a_N,b_N}\right) \left(E^{(1)}_{g_1,h_1} \otimes E^{(2)}_{g_2,h_2} \otimes \cdots \otimes E^{(N)}_{g_N,h_N}\right)\right]$$

$$= Trace\left[\left(E^{(1)\dagger}_{a_1,b_1} E^{(1)}_{g_1,h_1}\right) \otimes \left(E^{(2)\dagger}_{a_2,b_2} E^{(2)}_{g_2,h_2}\right) \otimes \cdots \otimes \left(E^{(N)\dagger}_{a_N,b_N} E^{(N)}_{g_N,h_N}\right)\right]$$

$$= Trace\left[E^{(1)\dagger}_{a_1,b_1} E^{(1)}_{g_1,h_1}\right] Trace\left[E^{(2)\dagger}_{a_2,b_2} E^{(2)}_{g_2,h_2}\right] \cdots Trace\left[E^{(N)\dagger}_{a_N,b_N} E^{(N)}_{g_N,h_N}\right]$$

$$= (d_1\, \delta_{a_1,g_1}\delta_{b_1,h_1})\, (d_2\, \delta_{a_2,g_2}\delta_{b_2,h_2}) \cdots (d_N\, \delta_{a_N,g_N}\delta_{b_N,h_N}) = d\, \delta_{a,g}\, \delta_{b,h},$$

where we used the map between the sets $\{a^{(k)}, b^{(k)}\} \rightarrow \{a^{\otimes}, b^{\otimes}\}$, and the fact that $d = \prod_{k=1}^{k=N} d_k$. Thus we conclude $\left\langle E^{\otimes}_{a,b}, E^{\otimes}_{g,h} \right\rangle = Trace\left[E^{\otimes\dagger}_{a,b}, E^{\otimes}_{g,h}\right] = \delta_{a,g}\delta_{b,h}$. The orthogonality of the $\left\{E^{\otimes}_{a,b}\right\}$ means we can expand $\rho^{\otimes}$ in terms of the $\left\{E^{\otimes}_{a,b}\right\}$, yielding $\rho^{\otimes} = \frac{1}{d} \sum_{a,b\in\{0,1,2,\cdots,d-1\}} \alpha_{a,b}\, E^{\otimes}_{a,b}$.

Another property of the $E^{\otimes}_{a,b}$ we shall need is the result of the product $E^{\otimes}_{g,h}\, E^{\otimes}_{a,b}\, E^{\otimes\dagger}_{g,h}$. Using equation X, and the tensor nature of $E^{\otimes}_{a,b}$, we have $E^{\otimes}_{g,h}\, E^{\otimes}_{a,b}\, E^{\otimes\dagger}_{g,h} =$

$$\left(E^{(1)}_{g_1,h_1} \otimes E^{(2)}_{g_2,h_2} \otimes \cdots \otimes E^{(N)}_{g_N,h_N}\right) \left(E^{(1)}_{a_1,b_1} \otimes E^{(2)}_{a_2,b_2} \otimes \cdots \otimes E^{(N)}_{a_N,b_N}\right) \left(E^{(1)}_{g_1,h_1} \otimes E^{(2)}_{g_2,h_2} \otimes \cdots \otimes E^{(N)}_{g_N,h_N}\right)^{\dagger}$$

$$= \left(E^{(1)}_{g_1,h_1} \otimes E^{(2)}_{g_2,h_2} \otimes \cdots \otimes E^{(N)}_{g_N,h_N}\right) \left(E^{(1)}_{a_1,b_1} \otimes E^{(2)}_{a_2,b_2} \otimes \cdots \otimes E^{(N)}_{a_N,b_N}\right) \left(E^{(1)\dagger}_{g_1,h_1} \otimes E^{(2)\dagger}_{g_2,h_2} \otimes \cdots \otimes E^{(N)\dagger}_{g_N,h_N}\right)$$

$$= \left(E^{(1)}_{g_1,h_1} E^{(1)}_{a_1,b_1} E^{(1)\dagger}_{g_1,h_1}\right) \otimes \left(E^{(2)}_{g_2,h_2} E^{(2)}_{a_2,b_2} E^{(2)\dagger}_{g_2,h_2}\right) \otimes \cdots \otimes \left(E^{(N)}_{g_N,h_N} E^{(N)}_{a_N,b_N} E^{(N)\dagger}_{g_N,h_N}\right)$$

$$= \left({\omega_1}^{a_1h_1-b_1g_1} E^{(1)}_{a_1,b_1}\right) \otimes \left({\omega_2}^{a_2h_2-b_2g_2} E^{(2)}_{a_2,b_2}\right) \otimes \cdots \otimes \left({\omega_N}^{a_Nh_N-b_Ng_N} E^{(N)}_{a_N,b_N}\right)$$

$$= \Omega^c E^{(1)}_{a_1,b_1} \otimes E^{(2)}_{a_2,b_2} \otimes \cdots \otimes E^{(N)}_{a_N,b_N} = \Omega^c E^{\otimes}_{a,b}, \tag{XIII}$$

where $\omega_k = e^{\frac{2\pi i}{d_k}}$, $\Omega = e^{\frac{2\pi i}{d}}$, and $c = \sum_{k=1}^{k=N} (a_k h_k - b_k g_k)\frac{d}{d_k}$.

XI.c The Channel $\mathcal{E}^{\otimes}$ is Unital and Diagonal in the $E^{\otimes}_{a,b}$ Basis

The channel $\mathcal{E}^{\otimes}$ is diagonal in the $E^{\otimes}_{a,b}$ basis. To see this, note that

$$\mathcal{E}^{\otimes}\left(E^{\otimes}_{a,b}\right) = \mathcal{E}^{\otimes}\left(E^{(1)}_{a_1,b_1} \otimes E^{(2)}_{a_2,b_2} \otimes \cdots \otimes E^{(N)}_{a_N,b_N}\right) \tag{XIV}$$

$$= \mathcal{E}^{(1)} \otimes \mathcal{E}^{(2)} \otimes \cdots \otimes \mathcal{E}^{(N)}\left(E^{(1)}_{a_1,b_1} \otimes E^{(2)}_{a_2,b_2} \otimes \cdots \otimes E^{(N)}_{a_N,b_N}\right)$$

$$= \mathcal{E}^{(1)}\left(E^{(1)}_{a_1,b_1}\right) \otimes \mathcal{E}^{(2)}\left(E^{(2)}_{a_2,b_2}\right) \otimes \cdots \otimes \mathcal{E}^{(N)}\left(E^{(N)}_{a_N,b_N}\right)$$

$$= \left(\alpha^{(1)}_{a_1,b_1} E^{(1)}_{a_1,b_1}\right) \otimes \left(\alpha^{(2)}_{a_2,b_2} E^{(2)}_{a_2,b_2}\right) \otimes \cdots \otimes \left(\alpha^{(N)}_{a_N,b_N} E^{(N)}_{a_N,b_N}\right)$$

$$= \alpha^{(1)}_{a_1,b_1} \alpha^{(2)}_{a_2,b_2} \cdots \alpha^{(N)}_{a_N,b_N} \left(E^{(1)}_{a_1,b_1}\right) \otimes \left(E^{(2)}_{a_2,b_2}\right) \otimes \cdots \otimes \left(E^{(N)}_{a_N,b_N}\right)$$

$$= \Lambda_{a,b}\left(E^{(1)}_{a_1,b_1}\right) \otimes \left(E^{(2)}_{a_2,b_2}\right) \otimes \cdots \otimes \left(E^{(N)}_{a_N,b_N}\right) = \Lambda_{a,b} E^{\otimes}_{a,b},$$

where $\Lambda_{a,b} = \alpha^{(1)}_{a_1,b_1} \alpha^{(2)}_{a_2,b_2} \cdots \alpha^{(N)}_{a_N,b_N}$, and we used the bijective map $\left\{a^{(k)}, b^{(k)}\right\} \Longleftrightarrow \{a^{\otimes}, b^{\otimes}\}$ to move back and forth between the operator basis set $\left\{E^{\otimes}_{a,b}\right\}$ and the operator basis set $\left\{E^{(1)}_{a_1,b_1} \otimes E^{(2)}_{a_2,b_2} \otimes \cdots \otimes E^{(N)}_{a_N,b_N}\right\}$. Thus the tensor product of diagonal qudit channels yields a diagonal qudit channel.

Next note that

$$E^{\otimes}_{0,0} = E^{(1)}_{0,0} \otimes E^{(2)}_{0,0} \otimes \cdots \otimes E^{(N)}_{0,0} = I_{d_1} \otimes I_{d_2} \otimes \cdots \otimes I_{d_N} = I_d.$$

Taking a special case of the result in equation XIV, we obtain

$$\mathcal{E}^{\otimes}(I_d) = \mathcal{E}^{\otimes}\left(E^{\otimes}_{0,0}\right) = \mathcal{E}^{(1)}\left(E^{(1)}_{0,0}\right) \otimes \mathcal{E}^{(2)}\left(E^{(2)}_{0,0}\right) \otimes \cdots \otimes \mathcal{E}^{(N)}\left(E^{(N)}_{0,0}\right)$$

$$= \mathcal{E}^{(1)}(I_{d_1}) \otimes \mathcal{E}^{(2)}(I_{d_2}) \otimes \cdots \otimes \mathcal{E}^{(N)}(I_{d_N}) = I_{d_1} \otimes I_{d_2} \otimes \cdots \otimes I_{d_N} = I_d.$$

We conclude that $\mathcal{E}^{\otimes}(I_d) = I_d$, and the channel $\mathcal{E}^{\otimes}$ is unital. Thus the tensor product of diagonal, unital qudit channels yields a diagonal unital qudit channel.

As an example, consider the product of two qubit (diagonal) unital channels, $\mathcal{E}^{(1)}$ with diagonal parameters $\{\lambda_1\,,\,\lambda_2\,,\,\lambda_3\}$, and $\mathcal{E}^{(2)}$ with diagonal parameters $\{\xi_1\,,\,\xi_2\,,\,\xi_3\}$. The product channel $\mathcal{E}^{\otimes} = \mathcal{E}^{(1)} \otimes \mathcal{E}^{(2)}$ is a diagonal, unital channel, taking an input vector of $(d_1\,d_2)^2 - 1 = 4^2 - 1 = 15$ input density matrix coefficients $\alpha_{a,b}$ to the output density matrix coefficients $\widetilde{\alpha}_{a,b}$, as shown below.

$$
\begin{array}{lcl}
\{\ basis\ element\ I_2 \otimes \sigma_x\ \} & \left[\begin{array}{l}\alpha_{0,1}\end{array}\right. & \left[\begin{array}{l}\widetilde{\alpha}_{0,1} = \xi_1\,\alpha_{0,1}\end{array}\right. \\
\{\ basis\ element\ I_2 \otimes \sigma_y\ \} & \alpha_{0,2} & \widetilde{\alpha}_{0,2} = \xi_2\,\alpha_{0,2} \\
\{\ basis\ element\ I_2 \otimes \sigma_z\ \} & \alpha_{0,3} & \widetilde{\alpha}_{0,3} = \xi_3\,\alpha_{0,3} \\
\{\ basis\ element\ \sigma_x \otimes I_2\ \} & \alpha_{1,0} & \widetilde{\alpha}_{1,0} = \lambda_1\ \alpha_{1,0} \\
\{\ basis\ element\ \sigma_y \otimes I_2\ \} & \alpha_{2,0} & \widetilde{\alpha}_{2,0} = \lambda_2\ \alpha_{2,0} \\
\{\ basis\ element\ \sigma_z \otimes I_2\ \} & \alpha_{3,0} & \widetilde{\alpha}_{3,0} = \lambda_3\ \alpha_{3,0} \\
\{\ basis\ element\ \sigma_x \otimes \sigma_x\ \} & \alpha_{1,1} & \widetilde{\alpha}_{1,1} = \lambda_1\,\xi_1\,\alpha_{1,1} \\
\{\ basis\ element\ \sigma_x \otimes \sigma_y\ \} & \alpha_{1,2} \quad \xrightarrow{\mathcal{E}} & \widetilde{\alpha}_{1,2} = \lambda_1\,\xi_2\,\alpha_{1,2} \\
\{\ basis\ element\ \sigma_x \otimes \sigma_z\ \} & \alpha_{1,3} & \widetilde{\alpha}_{1,3} = \lambda_1\,\xi_3\,\alpha_{1,3} \\
\{\ basis\ element\ \sigma_y \otimes \sigma_x\ \} & \alpha_{2,1} & \widetilde{\alpha}_{2,1} = \lambda_2\,\xi_1\,\alpha_{2,1} \\
\{\ basis\ element\ \sigma_y \otimes \sigma_y\ \} & \alpha_{2,2} & \widetilde{\alpha}_{2,2} = \lambda_2\,\xi_2\,\alpha_{2,2} \\
\{\ basis\ element\ \sigma_y \otimes \sigma_z\ \} & \alpha_{2,3} & \widetilde{\alpha}_{2,3} = \lambda_2\,\xi_3\,\alpha_{2,3} \\
\{\ basis\ element\ \sigma_z \otimes \sigma_x\ \} & \alpha_{3,1} & \widetilde{\alpha}_{3,1} = \lambda_3\,\xi_1\,\alpha_{3,1} \\
\{\ basis\ element\ \sigma_z \otimes \sigma_y\ \} & \alpha_{3,2} & \widetilde{\alpha}_{3,2} = \lambda_3\,\xi_2\,\alpha_{3,2} \\
\{\ basis\ element\ \sigma_z \otimes \sigma_z\ \} & \left.\begin{array}{l}\alpha_{3,3}\end{array}\right] & \left.\begin{array}{l}\widetilde{\alpha}_{3,3} = \lambda_3\,\xi_3\,\alpha_{3,3}\end{array}\right]
\end{array}
$$

XI.d The Average Output State $\widetilde{\Phi}$ of an Optimal Ensemble for $\mathcal{E}^{\otimes}$ is $\propto I_d$

Define the set of d^3 operators $\{F^{\otimes}_{a,b,c}\}$ as $F^{\otimes}_{a,b,c} = e^{\frac{2\pi i c}{d}}\,E^{\otimes}_{a,b}$. Using the bijective map between the $\left\{a_i^{(k)}, b_i^{(k)}\right\}$ and the $\{a^{\otimes}, b^{\otimes}\}$, we expand the $F^{\otimes}_{a,b,c}$ in terms of a phase $e^{\frac{2\pi i}{d}}$ and

the $\left\{E^{(k)}_{a_k,b_k}\right\}$. The expression for $F^{\otimes}_{a,b,c}$ becomes

$$F^{\otimes}_{a,b,c} = e^{\frac{2\pi ic}{d}} E^{\otimes}_{a,b} = e^{\frac{2\pi ic}{d}} E^{(1)}_{a_1,b_1} \otimes E^{(2)}_{a_2,b_2} \otimes \cdots \otimes E^{(N)}_{a_N,b_N}.$$

The set of operators $\{F^{\otimes}_{a,b,c}\}$ are the product of a phase $e^{\frac{2\pi ic}{d}}$ and the tensor products of the individual operators $\left\{E^{(k)}_{a_k,b_k}\right\}$. The $\{F^{\otimes}_{a,b,c}\}$ are unitary operators, inheriting this behavior from the unitary nature of the phase factor and the unitary nature of the subsystem operators $\left\{E^{(k)}_{a_k,b_k}\right\}$. To see this, note

$$F^{\otimes\dagger}_{a,b,c} F^{\otimes}_{a,b,c} = \left(e^{\frac{2\pi ic}{d}} E^{(1)}_{a_1,b_1} \otimes E^{(2)}_{a_2,b_2} \otimes \cdots \otimes E^{(N)}_{a_N,b_N}\right)^{\dagger} \left(e^{\frac{2\pi ic}{d}} E^{(1)}_{a_1,b_1} \otimes E^{(2)}_{a_2,b_2} \otimes \cdots \otimes E^{(N)}_{a_N,b_N}\right)$$

$$= e^{\frac{-2\pi ic}{d}} e^{\frac{2\pi ic}{d}} \left(E^{(1)\dagger}_{a_1,b_1} \otimes E^{(2)\dagger}_{a_2,b_2} \otimes \cdots \otimes E^{(N)\dagger}_{a_N,b_N}\right) \left(E^{(1)}_{a_1,b_1} \otimes E^{(2)}_{a_2,b_2} \otimes \cdots \otimes E^{(N)}_{a_N,b_N}\right)$$

$$= 1 \left(E^{(1)\dagger}_{a_1,b_1} E^{(1)}_{a_1,b_1}\right) \otimes \left(E^{(2)\dagger}_{a_2,b_2} E^{(2)}_{a_2,b_2}\right) \otimes \cdots \otimes \left(E^{(N)\dagger}_{a_N,b_N} E^{(N)}_{a_N,b_N}\right) = I_{d_1} \otimes I_{d_2} \otimes \cdots \otimes I_{d_N} = I_d,$$

where we used the unitary nature of the $\left\{E^{(k)}_{a_k,b_k}\right\}$ to say $E^{(k)\dagger}_{a_k,b_k} E^{(k)}_{a_k,b_k} = I_{d_k}$.

The $\{F^{\otimes}_{a,b,c}\}$ form a finite group which we shall call $\mathcal{Q}$. Consider a $d = \prod_{k=1}^{N} d_k$ dimensional representation Γ of $\mathcal{Q}$, built from the tensor products of the N d_k-dimensional matrix representations for $\hat{X}$ and $\hat{Z}$. To see why the representation Γ is irreducible, recall the relation for irreducibility from [27] discussed above. A necessary and sufficient condition for a representation Γ of a finite group $\mathcal{Q}$ to be irreducible is if the relation $\frac{1}{\|\mathcal{Q}\|} \sum_{q\in\mathcal{Q}} \left| Trace\left[\Gamma(q)\right] \right|^2 = 1$ is true[27]. Here $\|\mathcal{Q}\|$ is the order of the group $\mathcal{Q}$. Let the group $\mathcal{Q}$ be the set $\{F^{\otimes}_{a,b,c}\}$, where $a, b, c \in \{0, 1, 2, \ldots, d-1\}$. $\mathcal{Q}$ is of order d^3 and hence finite. Previously, we noted that $E^{\otimes}_{0,0} = I_d$ and $Trace\left[E^{\otimes\dagger}_{a,b}, E^{\otimes}_{g,h}\right] = \delta_{a,g}\delta_{b,h}$. Thus $Trace\left[E^{\otimes}_{a,b}\right] = d\,\delta_{a,0}\,\delta_{b,0}$. Computing the Trace sum yields

$$\frac{1}{\|\mathcal{Q}\|} \sum_{q\in\mathcal{Q}} \left| Trace\left[\Gamma(q)\right] \right|^2 = \frac{1}{d^3} \sum_{c\in\{0,1,2,\ldots,d-1\}} \sum_{b\in\{0,1,2,\ldots,d-1\}} \sum_{a\in\{0,1,2,\ldots,d-1\}} \left| Trace[F^{\otimes}_{a,b,c}] \right|^2$$

$$= \frac{1}{d^3} \sum_{c\in\{0,1,2,\ldots,d-1\}} \sum_{b\in\{0,1,2,\ldots,d-1\}} \sum_{a\in\{0,1,2,\ldots,d-1\}} \left| Trace[e^{\frac{2\pi i}{d}} E^{\otimes}_{a,b}] \right|^2$$

$$= \frac{1}{d^3} \sum_{c\in\{0,1,2,\ldots,d-1\}} \sum_{b\in\{0,1,2,\ldots,d-1\}} \sum_{a\in\{0,1,2,\ldots,d-1\}} \left| e^{\frac{2\pi i}{d}} Trace[E^{\otimes}_{a,b}] \right|^2$$

$$= \frac{1}{d^3} \sum_{c\in\{0,1,2,\ldots,d-1\}} \sum_{b\in\{0,1,2,\ldots,d-1\}} \sum_{a\in\{0,1,2,\ldots,d-1\}} \left| Trace[E^{\otimes}_{a,b}] \right|^2$$

$$= \frac{1}{d^3}\, d \sum_{b\in\{0,1,2,\ldots,d-1\}} \sum_{a\in\{0,1,2,\ldots,d-1\}} \left| \, d\, \delta_{a,0}\, \delta_{b,0} \, \right|^2 \; = \; 1.$$

Thus we find the representation Γ of the group $\mathcal{Q}$ is irreducible.

The fact that the channel $\mathcal{E}^{\otimes}$ is diagonal in the operator basis $\left\{E^{\otimes}_{a,b}\right\}$, coupled with the equation XIII result that $E^{\otimes}_{g,h}\, E^{\otimes}_{a,b}\, E^{\otimes\dagger}_{g,h} \; = \; \Omega^c\, E^{\otimes}_{a,b}$, and the equation XIV result that $\mathcal{E}\left(E^{\otimes}_{a,b}\right) \; = \; \Lambda_{a,b}\, E^{\otimes}_{a,b}$, allows us to conclude the operators $\left\{F^{\otimes}_{a,b,c}\right\}$ and the channel $\mathcal{E}^{\otimes}$ commute.

$$F^{\otimes}_{g,h,j}\, \mathcal{E}\left(\rho\right)\, F^{\otimes\dagger}_{g,h,j} \; = \; E^{\otimes}_{g,h}\, \mathcal{E}\left(\rho\right)\, E^{\otimes\dagger}_{g,h} \; = \; E^{\otimes}_{g,h} \left(\frac{1}{d} \sum_{a,b} \alpha_{a,b}\, \Lambda_{a,b}\, E^{\otimes}_{a,b} \right) E^{\otimes\dagger}_{g,h} \qquad \text{(XV)}$$

$$= \frac{1}{d} \sum_{a,b} \alpha_{a,b}\, \Lambda_{a,b}\, E^{\otimes}_{g,h}\, E^{\otimes}_{a,b}\, E^{\otimes\dagger}_{g,h} \; = \; \mathcal{E}\left(E^{\otimes}_{g,h}\, \rho E^{\otimes\dagger}_{g,h}\right) \; = \; \mathcal{E}\left(F^{\otimes}_{g,h,j}\, \rho F^{\otimes\dagger}_{g,h,j}\right).$$

Note that the product channel analysis in equation XV is essentially the same derivation as was done in equation XI for qudits in the $\hat{X}^a\, \hat{Z}^b$ operator basis, where the representation for the group $\left\{F^{\otimes}_{a,b,c}\right\}$ is built from the d_k-dimensional matrices for $\hat{X}$ and $\hat{Z}$ using the basis association in Section XI.a. [9]

This is the key criterion for ensemble achievability. Since the $\left\{F^{\otimes}_{a,b,c}\right\}$ are unitary, $F^{\otimes}_{g,h,j}\, \rho F^{\otimes\dagger}_{g,h,j}$ is a valid density operator. Applying any member of $\{F^{\otimes}_{a,b,c}\}$ to an output optimal ensemble $\{p_i\,,\, \widetilde{\rho}^{\otimes}_i\}$ yields an achievable ensemble. Since the group representation we are using for $\{F^{\otimes}_{a,b,c}\}$ is irreducible, we can apply Schur's Lemma and conclude the average output state $\widetilde{\Phi}^{\otimes}$ for an optimal ensemble for the product channel $\mathcal{E}^{\otimes}$ must equal $\frac{1}{d}\, I_d$.

The remainder of the analysis for diagonal unital qudit channels uses the Schumacher and Westmoreland results summarized in equations V, VI, VII and VIII in the manner seen

[9] See Section XII for $d = 3$ examples of the matrices for $\hat{X}$ and $\hat{Z}$.

previously, and directly carries over to the product channel case. Thus we conclude for the product channel $\mathcal{E}^{\otimes}$, the HSW channel capacity is

$$\mathcal{C}_1 \;=\; \log_2(d) \;-\; min_\rho\, \mathcal{S}\left(\mathcal{E}^{\otimes}(\rho)\right) \;=\; \sum_{k=1}^{N} \log_2(d_k) \;-\; min_\rho\, \mathcal{S}\left(\mathcal{E}^{\otimes}(\rho)\right).$$

XII Non-Diagonal Qudit Unital Channels

Consider the following channel. [10]

$$\mathcal{E}(\rho) \;=\; \frac{1 \,-\, \alpha \,-\, \beta}{d}\, Trace[\rho]\; I_d \;+\; \alpha\, \rho \;+\; \beta\, \rho^T \;.$$

The constants α and β are real numbers. Here ρ^T denotes the transpose of ρ, not the conjugate transpose. The map is trace preserving since

$$Trace[\mathcal{E}(\rho)] \;=\; \frac{1 \,-\, \alpha \,-\, \beta}{d}\, Trace[\rho]\; Trace[I_d] \;+\; \alpha\, Trace[\rho] \;+\; \beta\, Trace[\rho^T]$$

$$=\; \frac{1 \,-\, \alpha \,-\, \beta}{d}\, Trace[\rho]\; Trace[I_d] \;+\; \alpha\, Trace[\rho] \;+\; \beta\, Trace[\rho]$$

$$=\; (\,1 \,-\, \alpha \,-\, \beta\,)\, Trace[\rho] \;+\; \alpha\, Trace[\rho] \;+\; \beta\, Trace[\rho] \;=\; Trace[\rho]\,.$$

The map $\mathcal{E}$ is also linear, since for complex constants μ and ν, and density operators ρ and ϕ, we have

$$\mathcal{E}\left(\mu\rho + \nu\phi\right) = \frac{1-\alpha-\beta}{d}\, Trace\left[\mu\rho + \nu\phi\right]\; I_d + \alpha\left(\mu\rho + \nu\phi\right) + \beta\left(\mu\rho + \nu\phi\right)^T$$

$$= \mu\frac{1-\alpha-\beta}{d}\, Trace[\rho]\, I_d + \nu\frac{1-\alpha-\beta}{d}\, Trace[\phi]\, I_d + \mu\alpha\rho + \nu\alpha\phi + \mu\beta\rho^T + \nu\beta\,\phi^T$$

$$= \mu\left(\frac{1-\alpha-\beta}{d}\, Trace[\rho]\, I_d + \alpha\rho + \beta\rho^T\right) + \nu\left(\frac{1-\alpha-\beta}{d}\, Trace[\phi]\, I_d + \alpha\phi + \beta\phi^T\right)$$

$$=\; \mu\,\mathcal{E}(\rho) \;+\; \nu\,\mathcal{E}(\phi)\,.$$

[10] Dr. Eric Rains made substantial contributions to the work in this section.

In addition, the channel $\mathcal{E}$ is unital, since

$$\mathcal{E}\,(I_d) = \frac{1-\alpha-\beta}{d}\,Trace[I_d]\,I_d + \alpha\,I_d + \beta\,I_d^T = (1-\alpha-\beta)\,I_d + \alpha\,I_d + \beta\,I_d = I_d.$$

Thus $\mathcal{E}$ is a linear, trace preserving, unital map.

Let us consider as a specific example the qutrit case. Choi's criterion [26] tells us that the qutrit map $\mathcal{E}$ will be completely positive if and only if the following three conditions on α and β are simultaneously met.[11]

$$8\,\alpha\,+\,2\,\beta\,\geq\,-1 \qquad and \qquad \alpha\,+\,4\,\beta\,\leq\,1 \qquad and \qquad \alpha\,-\,2\,\beta\,\leq\,1\,.$$

For example, the values $\alpha\,=\,\beta\,=\,\frac{1}{5}$ yield in respective order, $3\,\geq\,-1$, $1\,\leq\,1$, and $-\frac{1}{5}\,\leq\,1$, indicating the qutrit map $\mathcal{E}$ is a completely positive map with these $\alpha\,,\,\beta$ values. Thus, for some set of α and β (E.g.: by construction in this case), the qutrit map $\mathcal{E}$ is a linear, trace preserving, completely positive, unital map.

The interesting fact for the qutrit $\mathcal{E}$ channel is that it is not a diagonal unital channel. To see this, note that for qutrits, the Generalized Pauli basis consists of the $3^2\;=\;9$ operators $\left\{\,\hat{I}_3\,,\,\hat{X}\,,\,\hat{X}^2\,,\,\hat{Z}\,,\,\hat{Z}^2\,,\,\hat{X}\,\hat{Z}\,,\,\hat{X}^2\,\hat{Z}\,,\,\hat{X}\,\hat{Z}^2\,,\,\hat{X}^2\,\hat{Z}^2\,\right\}$. A three dimensional ($d=3$) representation for the qutrit operators $\hat{X}$ and $\hat{Z}$ are the 3 by 3 matrices:

$$\hat{X}\;=\;\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \qquad and \qquad \hat{Z}\;=\;\begin{bmatrix} 1 & 0 & 0 \\ 0 & e^{\frac{2\,\pi\,i}{3}} & 0 \\ 0 & 0 & e^{\frac{4\,\pi\,i}{3}} \end{bmatrix}.$$

Note $\hat{X}^3\;=\;\hat{Z}^3\;=\;I_3$. The transpose operation acting on $\hat{X}$ yields $\hat{X}^T\;=\;\hat{X}^2$ and acting

[11] Please see appendix C for details of this calculation.

on $\hat{Z}$ yields $\hat{Z}^T = \hat{Z}$. In this basis, we represent ρ as shown below.

$$\rho \equiv \begin{bmatrix} \Gamma_I \\ \Gamma_X \\ \Gamma_{X^2} \\ \Gamma_Z \\ \Gamma_{Z^2} \\ \Gamma_{XZ} \\ \Gamma_{X^2Z} \\ \Gamma_{XZ^2} \\ \Gamma_{X^2Z^2} \end{bmatrix} \qquad where \qquad \rho = \frac{1}{3}\left(I_3 + \sum_{(a,b)\in\Upsilon} \Gamma_{a,b}\, \hat{X}^a\, \hat{Z}^b \right) .$$

(Please see appendix B for the definition of Υ.) The action of the channel in this basis is:

$$\mathcal{E}(\rho) \equiv \begin{bmatrix} \acute{\Gamma}_I \\ \acute{\Gamma}_X \\ \acute{\Gamma}_{X^2} \\ \acute{\Gamma}_Z \\ \acute{\Gamma}_{Z^2} \\ \acute{\Gamma}_{XZ} \\ \acute{\Gamma}_{X^2Z} \\ \acute{\Gamma}_{XZ^2} \\ \acute{\Gamma}_{X^2Z^2} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \alpha & \beta & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \beta & \alpha & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \alpha+\beta & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \alpha+\beta & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \alpha & \beta\Omega & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \beta\Omega^2 & \alpha & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \alpha & \beta\Omega^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \beta\Omega & \alpha \end{bmatrix} \begin{bmatrix} \Gamma_I \\ \Gamma_X \\ \Gamma_{X^2} \\ \Gamma_Z \\ \Gamma_{Z^2} \\ \Gamma_{XZ} \\ \Gamma_{X^2Z} \\ \Gamma_{XZ^2} \\ \Gamma_{X^2Z^2} \end{bmatrix} .$$

As defined in appendix B, $\Omega = e^{\frac{2\pi i}{3}}$. Thus in the Generalized Pauli basis, the qutrit channel

$$\mathcal{E}(\rho) = \frac{1-\alpha-\beta}{3}\, Trace[\rho]\, I_3 + \alpha\,\rho + \beta\,\rho^T$$

is a unital channel which is *not* diagonal. We now show that the technique previously introduced for diagonal unital channels works for this channel, yielding an example of a *non-diagonal* unital qudit channel for which the HSW channel capacity is

$$\mathcal{C}_1 = \log_2(d) - min_\rho\, \mathcal{S}\left(\mathcal{E}^{\otimes}(\rho)\right) .$$

Consider the set of all d by d real orthogonal matrices. These matrices form a d dimensional representation of the orthogonal group $\mathcal{O}$[30]. These real orthogonal matrices O satisfy $O\,O^T \;=\; O^T\,O \;=\; I_d$, where the superscript T stands for the transpose operation. Consider the following.

$$\mathcal{E}\left(O\,\rho\,O^T\right) \;=\; \frac{1\,-\,\alpha\,-\,\beta}{d}\,Trace[\,O\,\rho\,O^T\,]\,I_d \;+\; \alpha\,O\,\rho\,O^T \;+\; \beta\,(O\,\rho\,O^T)^T$$

$$=\; \frac{1\,-\,\alpha\,-\,\beta}{d}\,Trace[\,\rho\,O^T\,O\,]\,I_d\,O\,O^T \;+\; \alpha\,O\,\rho\,O^T \;+\; \beta\,O\,\rho^T\,O^T$$

$$=\; \frac{1\,-\,\alpha\,-\,\beta}{d}\,Trace[\,\rho\,]\;O\,I_d\,O^T \;+\; \alpha\,O\,\rho\,O^T \;+\; \beta\,O\,\rho^T\,O^T$$

$$=\; O\,\left(\,\frac{1\,-\,\alpha\,-\,\beta}{d}\,Trace[\,\rho\,]\,I_d \;+\; \alpha\,\rho \;+\; \beta\,\rho^T\,\right)\,O^T \;=\; O\,\mathcal{E}(\rho)\,O^T\;.$$

We have used the cyclic nature of the trace to say $Trace[\,O\,\rho\,O^T\,] \;=\; Trace[\,\rho\,O^T\,O\,] \;=\; Trace[\,\rho\,]$ and the fact that $O\,O^T \;=\; O^T\,O \;=\; I_d$, so that $O^T \;=\; O^T\,O^T\,O$. This last relation leads to $\left(O^T\right)^T \;=\; \left(O^T\,O^T\,O\right)^T \;=\; O^T\,O\,O \;=\; O$.

The d dimensional real orthogonal matrices O are unitary, since $O^T \;=\; O^\dagger$. Thus if ρ is a valid density operator, so is $O\,\rho\,O^T \;=\; O\,\rho\,O^\dagger$. Consider an optimal ensemble $\{p_i\,,\,\rho_i\}$. Proceeding as before, since the O are unitary and $\mathcal{E}\left(O\,\rho_i\,O^T\right) \;=\; O\;\mathcal{E}(\rho_i)\;O^T$, we conclude the ensemble $\{\,p_i\,,\,O\,\rho_i\,O^T\,\}$ is also an optimal input ensemble. This leads us to conclude, given $\{p_i\,,\,\rho_i\}$is an optimal input ensemble, that all output ensembles of the form $\{\,p_i\,,\,O\,\mathcal{E}(\rho_i)\,O^T\,\}$ are achievable. By the uniqueness argument, we conclude that $O\,\widetilde{\Phi}\,O^T \;=\; \widetilde{\Phi} \quad \forall\; O \;\in \mathcal{O}$.

In order to invoke Schur's Lemma to complete the derivation and conclude that $\widetilde{\Phi} \;=\; \frac{1}{d}\,I_d$, we need to determine the validity of Schur's Lemma for the d by d orthogonal matrix representation of the orthogonal group $\mathcal{O}$. Instead of determining the irreducibility via the trace sum formula, we note the orthogonal group $\mathcal{O}$ is a compact Lie group[27]. Schur's Lemma is valid for any representation of a compact Lie group[31]. Thus we conclude $\widetilde{\Phi} \;=\; \frac{1}{d}\,I_d$. Continuing along the same line of reasoning as previously shown for diagonal

unital qudit channels, we obtain the HSW channel capacity for this channel as

$$\mathcal{C}_1 \;=\; \log_2(d) \;-\; min_\rho \mathcal{S}\left(\mathcal{E}^{\otimes}(\rho)\right) \;=\; \log_2(3) \;-\; min_\rho \mathcal{S}\left(\mathcal{E}^{\otimes}(\rho)\right) \;.$$

It should be noted that Prof. A. S. Holevo independently derived a similar channel example in [32].

XIII Channel Additivity and Minimum Channel Output Entropy for Unital, Diagonal Qudit Channels

The goal of this work was to analyze a simple class of qudit channels that were a natural extension to qubit unital channels, and prove that these channels exhibited strict channel capacity additivity, as the qubit unital channels do[19]. To that end, we hoped to show that the minimal output entropy of the tensor product of two unital diagonal channels was the sum of the minimal output entropies of the channels taken individually, thereby implying the the HSW channel capacity of the tensor product of qudit unital diagonal channels was strictly additive. Unfortunately, we were unsuccessful in our attempt to derive analytically the relationship between the minimal output entropy of the tensor product of the two channels with respect to the individual minimal output entropies. However, we can make the following connection between $\mathcal{C}_1$ and the minimum output entropy for diagonal, unital qudit channels.

Theorem: The Holevo-Schumacher-Westmoreland channel capacity of the tensor product of two diagonal unital qudit channels $\mathcal{A}$ and $\mathcal{B}$ is additive if and only if the minimum output entropy of the tensor product channel $\mathcal{A} \otimes \mathcal{B}$ is the sum of the minimum output entropy of the two channels $\mathcal{A}$ and $\mathcal{B}$ taken separately.

Proof:

Let $d_{\mathcal{A}}$ be the dimension of the qudit input to the unital, diagonal channel $\mathcal{A}$ and $d_{\mathcal{B}}$ be the dimension of the qudit input to unital, diagonal channel $\mathcal{B}$. From Section XI, we know that the tensor product channel $\mathcal{A} \otimes \mathcal{B}$ is a unital, diagonal channel. The HSW channel capacities of the three channels of interest are:

$$\mathcal{C}_1^{\mathcal{A}\otimes\mathcal{B}} \;=\; \log_2(d_{\mathcal{A}} d_{\mathcal{B}}) \;-\; min_{\rho_{\mathcal{A}\mathcal{B}}}\, \mathcal{S}\left(\mathcal{E}(\rho_{\mathcal{A}\mathcal{B}})\right),$$

$$\mathcal{C}_1^{\mathcal{A}} \;=\; \log_2(d_{\mathcal{A}}) \;-\; min_{\rho_{\mathcal{A}}}\, \mathcal{S}\left(\mathcal{E}(\rho_{\mathcal{A}})\right),$$

and

$$\mathcal{C}_1^{\mathcal{B}} \;=\; \log_2(d_{\mathcal{B}}) \;-\; min_{\rho_{\mathcal{B}}}\, \mathcal{S}\left(\mathcal{E}(\rho_{\mathcal{B}})\right).$$

Note that

$$min_{\rho_{\mathcal{A}\mathcal{B}}}\, \mathcal{S}\left(\mathcal{E}(\rho_{\mathcal{A}\mathcal{B}})\right) \;\leq\; min_{\rho_{\mathcal{A}}}\, \mathcal{S}\left(\mathcal{E}(\rho_{\mathcal{A}})\right) \;+\; min_{\rho_{\mathcal{B}}}\, \mathcal{S}\left(\mathcal{E}(\rho_{\mathcal{B}})\right), \tag{XVI}$$

since if $\rho_{\mathcal{A}}^O$ satisfies the minimum for channel $\mathcal{A}$ and $\rho_{\mathcal{B}}^O$ satisfies the minimum for channel $\mathcal{B}$, then the state $\rho_{\mathcal{A}\mathcal{B}} = \rho_{\mathcal{A}}^O \otimes \rho_{\mathcal{B}}^O$ has a von Neumann entropy at the $\mathcal{A} \otimes \mathcal{B}$ channel output of

$$\mathcal{S}\left(\mathcal{E}\left(\rho_{\mathcal{A}}^O \otimes \rho_{\mathcal{B}}^O\right)\right) = \mathcal{S}\left(\mathcal{E}\left(\rho_{\mathcal{A}}^O\right)\right) + \mathcal{S}\left(\mathcal{E}\left(\rho_{\mathcal{B}}^O\right)\right) = min_{\rho_{\mathcal{A}}}\, \mathcal{S}\left(\mathcal{E}\left(\rho_{\mathcal{A}}\right)\right) + min_{\rho_{\mathcal{B}}}\, \mathcal{S}\left(\mathcal{E}\left(\rho_{\mathcal{B}}\right)\right),$$

thereby implying $min_{\rho_{\mathcal{A}\mathcal{B}}}\, \mathcal{S}\left(\mathcal{E}(\rho_{\mathcal{A}\mathcal{B}})\right) \;\leq\; min_{\rho_{\mathcal{A}}}\, \mathcal{S}\left(\mathcal{E}(\rho_{\mathcal{A}})\right) \;+\; min_{\rho_{\mathcal{B}}}\, \mathcal{S}\left(\mathcal{E}(\rho_{\mathcal{B}})\right)$.

The inequality relation in equation XVI implies that

$$\mathcal{C}_1^{\mathcal{A}\otimes\mathcal{B}} \;\geq\; \mathcal{C}_1^{\mathcal{A}} \;+\; \mathcal{C}_1^{\mathcal{B}}.$$

Thus the condition for strict HSW channel additivity,

$$\mathcal{C}_1^{\mathcal{A}\otimes\mathcal{B}} \;=\; \mathcal{C}_1^{\mathcal{A}} \;+\; \mathcal{C}_1^{\mathcal{B}},$$

holds if and only if

$$min_{\rho_{\mathcal{AB}}} \mathcal{S}\left(\mathcal{E}(\rho_{\mathcal{AB}})\right) \;=\; min_{\rho_{\mathcal{A}}} \mathcal{S}\left(\mathcal{E}(\rho_{\mathcal{A}})\right) \;+\; min_{\rho_{\mathcal{B}}} \mathcal{S}\left(\mathcal{E}(\rho_{\mathcal{B}})\right).$$

$\triangle$ - *End of Proof.*

Thus, there exists a class of channels, the tensor products of unital, diagonal qudit channels, for which the HSW channel capacity is strictly additive.

XIV Discussion

The HSW channel capacity for single qubit unital channels was originally derived in [25] as

$$\mathcal{C}_1 \;=\; 1 \;-\; min_{\rho} \mathcal{S}\left(\mathcal{E}(\rho)\right).$$

This result was extended in [19] to the tensor product of single qubit unital channels. Our method for deriving the HSW channel capacity depends on the qudit unital channel being diagonal,[12] so our method only allows us to conclude that

$$\mathcal{C}_1 \;=\; \log_2(d) \;-\; min_{\rho} \mathcal{S}\left(\mathcal{E}(\rho)\right)$$

holds for diagonal unital channels. (Recall that all qubit unital channels are diagonal in some characteristic input and output basis.)

Our proof was handcrafted in two key respects. The first was the choice of a fixed operator basis, the Generalized Pauli basis, in which the density matrix expansions were made. There exists the possibility that, given a specific channel, a custom operator basis could be constructed in which the channel $\mathcal{E}$ would be diagonal. This in essence is how the proof showing any unital qubit channel is diagonal in some operator basis was done in [25]. The

[12] With the exception of the example in Section XII.

second assumption was the explicit manner by which we showed ensemble achievability.

To summarize, we showed an output ensemble was achievable by

1) using a preordained unitary operator basis consisting of elements $g \in \mathcal{G}$

and

2) considering only diagonal channels in the basis $\mathcal{G}$.

The result was an algorithm by which we were able to determine, given an optimal ensemble $\{p_i\,,\,\rho_i\}$, if the output ensemble $\left\{p_i, g\tilde{\rho}_i g^{-1}\right\}$ was achievable for $g \in \mathcal{G}$.

The possibility remains that we could extend the technique developed in this chapter for diagonal unital channels to non-diagonal unital channels. This is what happened for the non-diagonal unital qutrit channel analyzed in section XII. It remains unclear how, for generic non-diagonal unital channels, finite unitary groups $\mathcal{G}$ with corresponding irreducible representations Γ, (such as in the case of the orthogonal group $\mathcal{O}$ in Section XII), can be found which have the behavior $g\,\mathcal{E}(\rho)\,g^\dagger \;=\; \mathcal{E}(\,g\,\rho\,g^\dagger\,) \quad \forall\, g \,\in\, \mathcal{G}$. It was this commutative behavior which in turn led to the achievability of all output ensembles generated by the elements of the group $\mathcal{G}$ acting on the output density matrices of an optimal input ensemble.

Indeed, it should be noted that the use of the *same* unitary group $\mathcal{G}$ acting on the input and output channel density matrices is not a requirement. One could consider two different unitary groups, $\mathcal{G}$ and $\mathcal{H}$, with corresponding representations $\Gamma(\mathcal{G})$ and $\Pi(\mathcal{H})$, such that for $g \,\in\, \mathcal{G}$ and $h \,\in\, \mathcal{H}$, $\Pi(h)\,\mathcal{E}(\rho)\,\Pi(h^\dagger) \;=\; \mathcal{E}\left(\Gamma(g)\,\rho\,\Gamma(g^\dagger)\right)$. The unitary nature of $\mathcal{G}$ ensures $\Gamma(g)\,\rho\,\Gamma(g^\dagger)$ is a valid density matrix, yielding achievability for all output ensembles generated through the application of $\mathcal{H}$. The unitary nature of the group $\mathcal{H}$ ensures $\Pi(h)\,\mathcal{E}(\rho)\,\Pi(h^\dagger)$ is a valid density matrix. In addition, the $\Pi(\mathcal{H})$ representation must be irreducible for the application of Schur's Lemma. The $\Gamma(\mathcal{G})$ representation need not be irreducible, since Schur's Lemma is applied only to output ensembles. (Note that we would always have $||\mathcal{H}|| \,\leq\, ||\mathcal{G}||$, should $\mathcal{G} \,\neq\, \mathcal{H}$.) A constructive procedure to find such unitary

groups $\mathcal{G}$ and $\mathcal{H}$, and corresponding representations $\Gamma(\mathcal{G})$ and $\Pi(\mathcal{H})$, or indicate no such groups $\mathcal{G}$ and $\mathcal{H}$ and/or representations exist, would be a useful extension to the analysis presented in this chapter.

In conclusion, we feel we have "overconstrained" the requirements for the proofs. As a result, we conjecture the relation

$$\mathcal{C}_1 \;=\; \log_2(d) \;-\; min_{\rho}\, \mathcal{S}\left(\mathcal{E}(\rho)\right)$$

holds for all unital qudit channels.

The diagonal unital qudit channel capacity result extends the connection between the minimum von Neumann entropy at the channel output and the HSW channel capacity, which had previously been established in the qubit case, to a non-empty set of channels in any dimension. This implies a more universal connection between the minimum von Neumann entropy at the channel output and the classical information capacity for that quantum channel than had previously been shown.

As a final remark, we note that the uniqueness result of Section VIII.b allows one to write $\mathcal{C}_1 \;=\; \kappa\left(\mathcal{E}\right) \;-\; \sum_i p_i\, \mathcal{S}\left(\mathcal{E}\left(\rho_i\right)\right)$ for any channel $\mathcal{E}$ and all optimum signalling ensembles $\{p_i\,,\, \rho_i\}$ for that channel. Thus the quantity $\sum_i p_i\, \mathcal{S}\left(\mathcal{E}\left(\rho_i\right)\right)$ is an invariant quantity across all optimal signalling ensembles for a given channel. This invariant quantity may be useful in future analyses of HSW channel capacity.

The author would also like to bring to the readers attention a paper by Professor A. S. Holevo which further discusses the techniques introduced in this chapter[32].

XV Appendix A: Donald's Equality

We prove Donald's Equality below[33]. Let ρ_i be a set of density matrices with a priori probabilities α_i, so that $\alpha_i \geq 0 \; \forall \; i$ and $\sum_i \alpha_i = 1$. Let ϕ be any density matrix, and define $\sigma = \sum_i \alpha_i \rho_i$. Then:

$$\sum_i \alpha_i \, D(\rho_i \| \phi) = D(\sigma \| \phi) + \sum_i \alpha_i \, D(\rho_i \| \sigma) \, .$$

Proof:

$$\sum_i \alpha_i \, D(\rho_i \| \phi) = \sum_i \alpha_i \, \{ Tr[\rho_i \log(\rho_i)] - Tr[\rho_i \log(\phi)] \}$$

$$= \sum_i \alpha_i \, \{ Tr[\rho_i \log(\rho_i)] \} - Tr[\sigma \log(\phi)]$$

$$= \{ Tr[\sigma \log(\sigma)] - Tr[\sigma \log(\sigma)] \} - Tr[\sigma \log(\phi)] + \sum_i \alpha_i \, Tr[\rho_i \log(\rho_i)]$$

$$= D(\sigma \| \phi) - Tr[\sigma \log(\sigma)] + \sum_i \alpha_i \, Tr[\rho_i \log(\rho_i)]$$

$$= D(\sigma \| \phi) + \sum_i \alpha_i \, \{ Tr[\rho_i \log(\rho_i)] - Tr[\rho_i \log(\sigma)] \}$$

$$= D(\sigma \| \phi) + \sum_i \alpha_i \, D(\rho_i \| \sigma) \, .$$

$\triangle$ - *End of Proof.*

XVI Appendix B: The Generalized Pauli Group

The Generalized Pauli operators $\hat{X}$ and $\hat{Z}$ are used in the qudit analysis. This section describes some of the properties of these operators. Their definitions are

$$\hat{X}|j\rangle \;=\; |j+1\,(mod\;d)\rangle \qquad and \qquad \hat{Z}|j\rangle \;=\; \Omega^{j}\,|j\rangle.$$

The quantity $\Omega \;=\; e^{\frac{2\pi i}{d}}$. Note that $\hat{X}^{d} \;=\; \hat{Z}^{d} \;=\; I_{d}$. The commutation relation of $\hat{X}$ and $\hat{Z}$ follows directly, yielding $\hat{Z}\hat{X} \;=\; \Omega\hat{X}\hat{Z}$. Using the fact that $\langle j+1|\hat{X}|j\rangle \;=\; 1$, taking the Hermitian conjugate of both sides yields $\langle j|\hat{X}^{\dagger}|j+1\rangle \;=\; 1$, allowing us to conclude $\hat{X}^{\dagger}|j\rangle \;=\; |j-1\,(mod\;d)\rangle$. This in turn implies $\hat{X}$ is unitary, since $\hat{X}\hat{X}^{\dagger} \;=\; \hat{X}^{\dagger}\hat{X} \;=\; I_{d}$. Similarly $\hat{Z}^{\dagger}|j\rangle \;=\; \Omega^{-j}\,|j\rangle$, from which it follows that $\hat{Z}$ is a unitary operator.

In our application of Schur's Lemma, we use the operator set of $E_{a,b} \;=\; \hat{X}^{a}\hat{Z}^{b}$, where $\{a,b\} \;=\; 0,1,2,\cdots,d-1$. We shall also use the operators $F_{a,b,c} \;=\; \Omega^{c}\hat{X}^{a}\hat{Z}^{b}$, where $\{a,b,c\} \;=\; 0,1,2,\cdots,d-1$. The operators $E_{a,b}$ and $F_{a,b,c}$ are unitary, since the composition of unitary operators is unitary. Note that $E^{\dagger}_{a,b} \;=\; \hat{Z}^{-b}\hat{X}^{-a}$ and $F^{\dagger}_{a,b,c} \;=\; \Omega^{-c}E^{\dagger}_{a,b}$.

We now show that any qudit density operator ρ can be expanded as

$$\rho \;=\; \frac{1}{d}\sum_{a,b\in\{0,1,2,\cdots,d-1\}} \alpha_{a,b}\;\hat{X}^{a}\;\hat{Z}^{b} \;=\; \frac{1}{d}\sum_{a,b\in\{0,1,2,\cdots,d-1\}} \alpha_{a,b}\;E_{a,b},$$

where the $\alpha_{a,b}$ are complex quantities. We shall work in the Hilbert-Schmidt operator norm, which for qudit operators A and B is defined as $\langle A,B\rangle \;=\; Trace[A^{\dagger}B]$. Define the rescaled operators $Q_{a,b} \;=\; \frac{E_{a,b}}{\sqrt{d}} \;=\; \frac{\hat{X}^{a}\hat{Z}^{b}}{\sqrt{d}}$. The operators $Q_{a,b}$ are a set of d^{2} orthonormal operators in the Hilbert-Schmidt inner product, as shown below.

$$\langle Q_{a,b}\,,\,Q_{q,r}\rangle \;=\; \frac{1}{d}\langle E_{a,b}\,,\,E_{q,r}\rangle \;=\; \frac{1}{d}\,Trace[E^{\dagger}_{a,b}\,E_{q,r}] \;=\; \frac{1}{d}\,Trace[\hat{Z}^{-b}\hat{X}^{-a}\,\hat{X}^{q}\hat{Z}^{r}] \quad \text{(XVII)}$$

$$(By\;the\;cyclic\;nature\;of\;trace) \;=\; \frac{1}{d}\,Trace[\hat{X}^{q-a}\,\hat{Z}^{r-b}] \;=\; \frac{1}{d}\sum_{j=0}^{d-1}\langle j|\hat{X}^{q-a}\,\hat{Z}^{r-b}|j\rangle$$

$$= \frac{1}{d}\sum_{j=0}^{d-1} \Omega^{(r-b)j} \langle j|\hat{X}^{q-a}|j\rangle = \frac{1}{d}\sum_{j=0}^{d-1} \Omega^{(r-b)j} \langle j|j+q-a\,(mod\,d)\rangle$$

$$= \frac{1}{d}\,\delta_{a,q}\sum_{j=0}^{d-1} \Omega^{(r-b)j} = \frac{1}{d}\,d\,\delta_{a,q}\,\delta_{b,r} = \delta_{a,q}\,\delta_{b,r}.$$

Here $\delta_{\alpha,\beta}$ is the Kronecker delta function. Recall any qudit density operator ρ can be written as

$$\rho = \sum_{a=0}^{d-1}\sum_{b=0}^{d-1} \beta_{a,b}\,|a\rangle\langle b|,$$

where the $\beta_{a,b}$ are complex quantities. We shall show that $|a\rangle\langle b|$ may be written as $|a\rangle\langle b| = \sum_{r=0}^{d-1}\sum_{s=0}^{d-1} \zeta_{r,s}\,Q_{r,s}$, where the $\zeta_{r,s}$ are complex quantities. Rescaling the $\zeta_{r,s}$, we will conclude that ρ may be written as

$$\rho = \sum_{a=0}^{d-1}\sum_{b=0}^{d-1} \alpha_{a,b}\,E_{a,b}.$$

To begin, write $Q_{r,s}$ as

$$Q_{r,s} = \frac{1}{\sqrt{d}}\sum_{j=0}^{d-1} \Omega^{js}\,|j+r\rangle\langle j|.$$

Define $\zeta_{r,s}$ as [34]

$$\zeta_{a,b} = Trace\Big[Q_{r,s}^{\dagger}|a\rangle\langle b|\Big] = \frac{1}{\sqrt{d}}\,Trace\left[\sum_{j=0}^{d-1} \Omega^{-js}\,|j\rangle\langle j+r|a\rangle\langle b|\right] \qquad \text{(XVIII)}$$

$$= \Big(Do\quad the\quad Trace\quad in\quad the\quad basis\quad \Big\{|i\rangle\Big\}\Big) \longrightarrow \frac{1}{\sqrt{d}}\sum_{i=0}^{d-1}\sum_{j=0}^{d-1} \Omega^{-js}\,\langle i|j\rangle\langle j+r|a\rangle\langle b|i\rangle$$

$$= \frac{1}{\sqrt{d}}\sum_{i=0}^{d-1}\sum_{j=0}^{d-1} \Omega^{-js}\,\delta_{b,i}\,\delta_{j+r,a}\,\delta_{i,j} = \frac{1}{\sqrt{d}}\sum_{j=0}^{d-1} \Omega^{-js}\,\delta_{j,b}\,\delta_{j+r,a} = \frac{1}{\sqrt{d}}\,\Omega^{-bs}\,\delta_{a,b+r},$$

where δ is the Kronecker delta function.

Consider the operator $L = |a\rangle\langle b|$, and the corresponding complex coefficients $\xi_{r,s} = \langle Q_{r,s}, L\rangle = Trace\Big[Q_{r,s}^{\dagger}|a\rangle\langle b|\Big]$. We would like to expand L as $L = \sum_{r,s}\,\langle Q_{r,s}\,,\,L\rangle\,Q_{r,s} = \sum_{r,s}\,\xi_{r,s}\,Q_{r,s}$. Note that $\|L\| = \sqrt{\langle L, L\rangle} = 1$. Using the result of equation XVIII, we can

conclude that

$$\sum_r \sum_s |\xi_{r,s}|^2 = \sum_r \sum_s \left| \frac{1}{\sqrt{d}} \Omega^{-bs} \delta_{a,b+r} \right|^2 = \frac{1}{d} \sum_r \sum_s |\delta_{a,b+r}|^2 = \frac{1}{d} d \sum_r |\delta_{a,b+r}|^2 = 1.$$

Thus $\sum_r \sum_s |\xi_{r,s}|^2 = 1 = \|L\|^2$. This fact for arbitrary a and b in $|a\rangle\langle b|$ allows us to conclude the $Q_{r,s}$ form a complete, orthonormal basis for the L's, and we can expand L in terms of the $Q_{r,s}$ $\forall$ a, b[35]. Thus the expansion $|a\rangle\langle b| = \sum_{r,s} \langle Q_{r,s}\,,\, L\rangle\, Q_{r,s}$ holds $\forall$ a, b. This leads to an expansion for the qudit density operator ρ.

$$\rho = \sum_{a=0}^{d-1} \sum_{b=0}^{d-1} \beta_{a,b} |a\rangle\langle b| = \sum_{a=0}^{d-1} \sum_{b=0}^{d-1} \beta_{a,b} \sum_{r=0}^{d-1} \sum_{s=0}^{d-1} \Big\langle Q_{r,s}, (|a\rangle\langle b|) \Big\rangle Q_{r,s} \tag{XIX}$$

$$= \sum_{r=0}^{d-1} \sum_{s=0}^{d-1} \sum_{a=0}^{d-1} \sum_{b=0}^{d-1} \beta_{a,b} \Big\langle Q_{r,s}, (|a\rangle\langle b|) \Big\rangle Q_{r,s} = \sum_{r=0}^{d-1} \sum_{s=0}^{d-1} \left\langle Q_{r,s}, \left(\sum_{a=0}^{d-1} \sum_{b=0}^{d-1} \beta_{a,b} |a\rangle\langle b| \right) \right\rangle Q_{r,s}$$

$$= \sum_{r=0}^{d-1} \sum_{s=0}^{d-1} \Big\langle Q_{r,s}, \rho \Big\rangle Q_{r,s} = \sum_{r=0}^{d-1} \sum_{s=0}^{d-1} \frac{\alpha_{r,s}}{\sqrt{d}} Q_{r,s} = \sum_{r=0}^{d-1} \sum_{s=0}^{d-1} \frac{\alpha_{r,s}}{\sqrt{d}} \frac{E_{r,s}}{\sqrt{d}} = \frac{1}{d} \sum_{r=0}^{d-1} \sum_{s=0}^{d-1} \alpha_{r,s} E_{r,s}$$

$$\textit{where} \quad \frac{\alpha_{r,s}}{\sqrt{d}} = \langle\, Q_{r,s}\,,\, \rho\,\rangle \quad \textit{or} \quad \textit{equivalently} \quad \alpha_{r,s} = \langle\, E_{r,s}\,,\, \rho\,\rangle.$$

The linearity of the inner product in the second argument was used to move the sum over the indices a and b inside the inner product.

To obtain the final form of the expansion for the qudit operator ρ we shall use, note that $E_{0,0} = I_d$. Our result above, $\langle E_{a,b}\,,\, E_{q,r}\rangle = Trace[E_{a,b}^\dagger E_{q,r}] = d\,\delta_{a,q}\,\delta_{b,r}$, tells us that $Trace(E_{a,b}) = d\,\delta_{a,0}\,\delta_{b,0}$. Thus of the d^2 possible $E_{a,b}$, only $E_{0,0}$ has nonzero $Trace$. The trace condition $Trace(\rho) = 1$ allows us to conclude $\alpha_{0,0} = 1$. Using this, let Υ denote the set of $d^2 - 1$ elements $a, b \in \{0, 1, 2, \cdots, d-1\}$ *with the exception that a and b cannot both be zero.* Then we may write the qudit density matrix ρ as $\rho = \frac{1}{d}\left(I_d + \sum_{(a,b)\in\Upsilon} \alpha_{a,b}\, E_{a,b}\right)$ with $\alpha_{a,b} = \langle\, E_{a,b}\,,\, \rho\,\rangle = Trace[E_{a,b}^\dagger \rho]$.

In the expansion of ρ above, there are $2d^2 - 2$ real, independent degrees of freedom in the set of coefficients $\alpha_{a,b}$. However, in the density operator ρ, there are only $d^2 - 1$ real, independent degrees of freedom. Hence there are constraint relations between the $\alpha_{a,b}$.

These constraints arise from the Hermitian nature of ρ. Note that $E_{a,b}^\dagger = \left(\hat{X}^a \hat{Z}^b\right)^\dagger = \hat{Z}^{-b}\hat{X}^{-a} = \Omega^{d-b)(d-a)}\hat{X}^{d-a}\hat{Z}^{d-b} = \Omega^{d-b)(d-a)}E_{d-a,d-b}$. Consideration of $\rho^\dagger = \rho$ then implies

$$\frac{1}{d}\left(I_d + \sum_{(a,b)\in\Upsilon} \alpha_{a,b}\, E_{a,b}\right) = \frac{1}{d}\left(I_d + \sum_{(a,b)\in\Upsilon} \alpha_{a,b}^*\, E_{a,b}^\dagger\right)$$

$$= \frac{1}{d}\left(I_d + \sum_{(a,b)\in\Upsilon} \alpha_{a,b}^*\, \Omega^{(d-a)(d-b)} E_{d-a,d-b}\right)$$

or $\alpha_{d-a,d-b} = \alpha_{a,b}^*\Omega^{(d-a)(d-b)}$. Here * indicates complex conjugation, and index arithmetic is modulo d.

For example, for qubits, $d = 2$, and $\Omega = e^{\frac{2\pi i}{2}} = e^{\pi i} = -1$. Applying the constraint equation above leads to $\alpha_{0,1}^*\Omega^{(2-0)(2-1)} = \alpha_{2-0,2-1}$ or $\alpha_{0,1}^* = \alpha_{0,1}$, implying the coefficient of $E_{0,1} = \hat{Z}$ must be real. Similarly, $\alpha_{1,0}^*\Omega^{(2-1)(2-0)} = \alpha_{2-1,2-0}$ or $\alpha_{1,0}^* = \alpha_{1,0}$, implying the coefficient of $E_{1,0} = \hat{X}$ must be real. Lastly, $\alpha_{1,1}^*\Omega^{(2-1)(2-1)} = \alpha_{2-1,2-1}$ or $-\alpha_{1,1}^* = \alpha_{1,1}$, implying the coefficient of $E_{1,1} = \hat{X}\hat{Z}$ must be pure imaginary. Note that $\hat{X} = \sigma_x$, $\hat{X}\hat{Z} = -i\sigma_y$, and $\hat{Z} = \sigma_z$. Hence we have reproduced the Bloch Sphere representation for qubits, $\rho = \frac{1}{2}\left(I_2 + \alpha_{1,0}\hat{X} + \alpha_{1,1}\hat{X}\hat{Z} + \alpha_{0,1}\hat{Z}\right) = \frac{1}{2}\left(I_2 + w_x\sigma_x + iw_y\left(-i\sigma_y\right) + w_z\sigma_z\right)$, with the w_k real. For qubits, we end up with $3 = d^2 - 1$ real independent parameters, and not $2d^2 - 2 = 6$. The constraint equations for the $\alpha_{a,b}$ eliminated three real degrees of freedom. In general, the constraint equations will eliminate $d^2 - 1$ real extra degrees of freedom, leaving $d^2 - 1$ actual real parameters.

XVII Appendix C: Choi's Criterion for Complete Positivity

In this appendix, we calculate the Choi matrix for the qutrit channel discussed in Section XII.

$$\mathcal{E}(\rho) \;=\; \frac{1 \,-\, \alpha \,-\, \beta}{d}\, Trace[\rho]\; I_d \;+\; \alpha\, \rho \;+\; \beta\, \rho^T \;.$$

The constants α and β are real numbers. Here ρ^T denotes the transpose of ρ, not the conjugate transpose.

The Choi matrix is, in this case, the 9 by 9 matrix with submatrix entries $\mathcal{E}(E_{ij})$, where $i, j \;=\; 1, 2, 3$. The E_{ij} are 3 by 3 matrices which have all zero entries except in the i'th row and the j'th column. The resulting Choi matrix is:

$$\begin{bmatrix} \mathcal{E}(E_{11}) & \mathcal{E}(E_{12}) & \mathcal{E}(E_{13}) \\ \mathcal{E}(E_{21}) & \mathcal{E}(E_{22}) & \mathcal{E}(E_{23}) \\ \mathcal{E}(E_{31}) & \mathcal{E}(E_{32}) & \mathcal{E}(E_{33}) \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1+\alpha+\beta}{3} & 0 & 0 & 0 & \alpha & 0 & 0 & 0 & \alpha \\ 0 & \frac{1-\alpha-\beta}{3} & 0 & \beta & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1-\alpha-\beta}{3} & 0 & 0 & 0 & \beta & 0 & 0 \\ 0 & \beta & 0 & \frac{1-\alpha-\beta}{3} & 0 & 0 & 0 & 0 & 0 \\ \alpha & 0 & 0 & 0 & \frac{1+\alpha+\beta}{3} & 0 & 0 & 0 & \alpha \\ 0 & 0 & 0 & 0 & 0 & \frac{1-\alpha-\beta}{3} & 0 & \beta & 0 \\ 0 & 0 & \beta & 0 & 0 & 0 & \frac{1-\alpha-\beta}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \beta & 0 & \frac{1-\alpha-\beta}{3} & 0 \\ \alpha & 0 & 0 & 0 & \alpha & 0 & 0 & 0 & \frac{1+\alpha+\beta}{3} \end{bmatrix} .$$

The quantum channel is a completely positive map if and only if the Choi matrix is positive semidefinite. Since the matrix is Hermitian by inspection, we need only find the eigenvalues of the matrix above, and constrain these eigenvalues to be ≥ 0 in order to conclude

the matrix is positive semi-definite[36]. The matrix eigenvalues are (with the eigenvalue multiplicity in parentheses after the eigenvalue):

$$1 + 8\alpha + 2\beta \ (one), \quad 1 - \alpha - 4\beta \ (three), \quad 1 - \alpha + 2\beta \ (five).$$

These eigenvalues lead to the following three constraints on α and β.

$$8\alpha + 2\beta \geq -1 \qquad and \qquad \alpha + 4\beta \leq 1 \qquad and \qquad \alpha - 2\beta \leq 1 \, .$$

If α and β are chosen to simultaneously satisfy these three constraints, then the resulting channel will be completely positive.

Chapter 2

Relative Entropy and Single Qubit Holevo-Schumacher-Westmoreland Channel Capacity

I Abstract

The relative entropy description of Holevo-Schumacher-Westmoreland (HSW) classical channel capacities is applied to single qubit quantum channels. A simple formula for the relative entropy of qubit density matrices in the Bloch sphere representation is derived. The formula is combined with the King-Ruskai-Szarek-Werner qubit channel ellipsoid picture [1] to analyze several unital and non-unital qubit channels in detail. An alternate proof is presented that the optimal HSW signalling states for single qubit unital channels are those states with minimal channel output entropy. The derivation is based on symmetries of the relative entropy formula, and the King-Ruskai-Szarek-Werner qubit channel ellipsoid picture[2].

II Introduction

In 1999, Benjamin Schumacher and Michael Westmoreland published a paper entitled *Optimal Signal Ensembles* [16] that elegantly described the classical (product state) channel capacity of quantum channels in terms of a function known as the relative entropy. Building upon this view, we study single qubit channels, adding the following two items to the

[1] See Section IX.a of Chapter 1.

[2] A preprint of this work can be found on the Los Alamos National Laboratory preprint server[37].

Schumacher-Westmoreland analysis.

I) A detailed understanding of the convex hull shape of the set of quantum states output by a channel. (The fact the set was convex has been known for some time, but the detailed nature of the convex geometry was unknown until recently.)

II) A useful mathematical representation (formula) for the relative entropy function, $\mathcal{D}(\,\rho\,\|\,\phi\,)$, when both ρ and ϕ are single qubit density matrices.

For single qubit channels, the work of King, Ruskai, Szarek, and Werner (KRSW) has provided a concise description of the convex hull set[25, 26]. In this chapter, we derive a useful formula for the relative entropy between qubit density matrices. Combining this formula with the KRSW convexity information, we present from a relative entropy perspective several results, some previously known, and others new, related to the (product state) classical channel capacity of quantum channels. These include:

I) In Chapter 1 we showed that the average output density matrix for *any* optimal set of signalling states that achieves the HSW classical channel capacity for a quantum channel is unique. This led to the result that for diagonal unital quantum channels such as single qubit unital channels, the average output ensemble density matrix must be $\frac{1}{d}\,\mathcal{I}_d$. This fact will allow us to see why only two orthogonal signalling states are needed to achieve the optimum classical channel capacity for single qubit unital channels, and why the a priori probabilities for these two signalling states are $\frac{1}{2}$, as previously shown in [25].

II) The single qubit relative entropy formula allows us to understand geometrically why the a priori probabilities for optimum signalling states for non-unital single qubit channels are not equal.

III) Examples of channels which require non-orthogonal signalling states to achieve optimal classical channel capacity are given. Such channels have been found before. Here these channels are presented in a geometrical fashion based on the relative entropy formula derived in Appendix A.

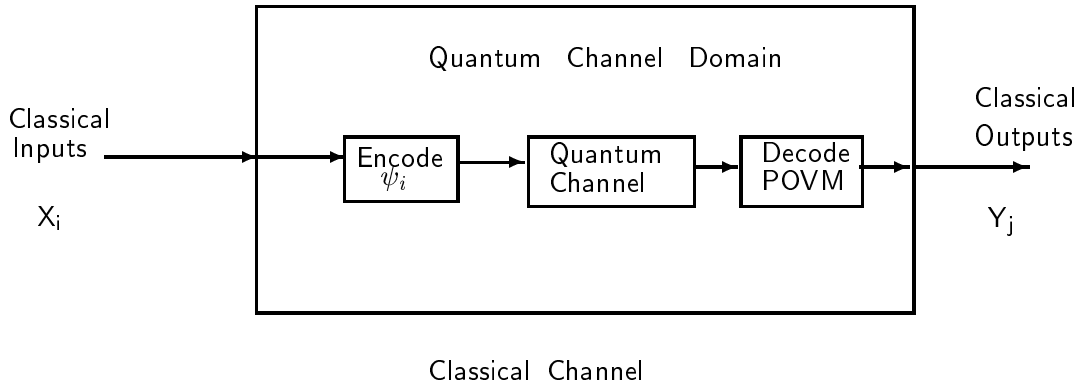

Figure 2.1: Transmission of classical information through a quantum channel.

III Background

III.a Classical Communication over Classical and Quantum Channels

This chapter discusses the transmission of classical information over quantum channels with no prior entanglement between the sender (Alice) and the recipient (Bob). In such a scenario, classical information is encoded into a set of quantum states ψ_i. These states are transmitted over a quantum channel. The perturbations encountered by the signals while transiting the channel are described using the Kraus representation formalism. A receiver at the channel output measures the perturbed quantum states using a POVM set. The resulting classical measurement outcomes represent the extraction of classical information from the channel output quantum states.

There are two common criteria for measuring the quality of the transmission of classical information over a channel, regardless of whether the channel is classical or quantum. These criteria are the (Product State) Channel Capacity[12, 13, 14] and the probability of error (Pe)[38]. In this chapter, we shall focus on the first criterion, the (Product State) Classical Information Capacity of a Quantum Channel, $\mathcal{C}_1$. This capacity, known as the Holevo-Schumacher-Westmoreland (HSW) classical channel capacity, was defined in Section IV of Chapter 1. As discussed in Sections VI and VII of Chapter 1,

the HSW channel capacity $\mathcal{C}_1$ can be expressed in terms of the relative entropy function $\mathcal{D}(\varrho \,\|\, \phi) = Tr[\varrho \log(\varrho) - \varrho \log(\phi)]$ as

$$\mathcal{C}_1 = \mathrm{Min}_\xi \qquad \mathrm{Max}_\varrho \qquad \mathcal{D}(\varrho \,\|\, \xi),$$

where $\varrho \in \mathcal{A} \equiv$ the set of possible channel output states for a channel $\mathcal{E}$ corresponding to pure state inputs, and $\xi \in$ the convex hull of $\mathcal{A}$.

IV Relative Entropy for Qubits in the Bloch Sphere Representation

The key formula we shall use extensively is the relative entropy in the Bloch sphere representation. Here ρ and ϕ have the respective Bloch sphere representations:[3]

$$\rho = \frac{1}{2}(\mathcal{I}_2 + \vec{\mathcal{W}} \bullet \vec{\sigma}) \qquad \textit{and} \qquad \phi = \frac{1}{2}(\mathcal{I}_2 + \vec{\mathcal{V}} \bullet \vec{\sigma}).$$

We define $\cos(\theta)$ as

$$\cos(\theta) = \frac{\vec{\mathcal{W}} \bullet \vec{\mathcal{V}}}{r\, q} \quad \textit{where} \quad r = \sqrt{\vec{\mathcal{W}} \bullet \vec{\mathcal{W}}} \;\; \textit{and} \;\; q = \sqrt{\vec{\mathcal{V}} \bullet \vec{\mathcal{V}}}\,.$$

In Appendix A, we prove the following formula for the relative entropy $\mathcal{D}(\varrho \,\|\, \psi)$ of two single qubit density matrices ϱ and ψ with Bloch sphere representations given above.

$$\mathcal{D}(\varrho \,\|\, \psi) = \frac{1}{2}\log_2\left(1 - r^2\right) + \frac{r}{2}\log_2\left(\frac{1+r}{1-r}\right) - \frac{1}{2}\log_2\left(1 - q^2\right) - \frac{\vec{\mathcal{W}} \bullet \vec{\mathcal{V}}}{2\, q}\log_2\left(\frac{1+q}{1-q}\right)$$

$$= \frac{1}{2}\log_2\left(1 - r^2\right) + \frac{r}{2}\log_2\left(\frac{1+r}{1-r}\right) - \frac{1}{2}\log_2\left(1 - q^2\right) - \frac{r\,\cos(\theta)}{2}\log_2\left(\frac{1+q}{1-q}\right),$$

where θ is the angle between $\vec{\mathcal{W}}$ and $\vec{\mathcal{V}}$, and r and q are as defined above.

When ϕ in $\mathcal{D}(\rho \,\|\, \phi)$ is the maximally mixed state $\phi = \frac{1}{2}\mathcal{I}_2$, we have $q = 0$, and $\mathcal{D}(\rho \,\|\, \phi)$

[3] Here $\mathcal{I}_2$ is the two by two identity matrix.

becomes the radially symmetric function

$$\mathcal{D}(\,\rho\,\|\,\phi\,) \;=\; \mathcal{D}\left(\,\rho\,\|\,\frac{1}{2}\,\mathcal{I}_2\,\right) \;=\; \frac{1}{2}\,\log_2\left(1-r^2\right) \;+\; \frac{r}{2}\,\log_2\left(\frac{1+r}{1-r}\right) \;=\; 1\;-\;\mathcal{S}(\,\rho\,)\;.$$

It is shown in Appendix A that $\mathcal{D}\left(\,\rho\,\|\,\frac{1}{2}\,\mathcal{I}_2\,\right) \;=\; 1\;-\;\mathcal{S}(\rho)$, where $\mathcal{S}(\rho)$ is the von Neumann entropy of ρ. In what follows, we shall often write $\mathcal{D}(\,\rho\,\|\,\phi\,)$ as $\mathcal{D}(\,\vec{\mathcal{W}}\,\|\,\vec{\mathcal{V}}\,)$, where $\vec{\mathcal{W}}$ and $\vec{\mathcal{V}}$ are the Bloch sphere vectors for ρ and ϕ respectively.

In what follows, we shall graphically determine the HSW channel capacity from the intersection of contours of constant relative entropy with the channel ellipsoid. To that end, and to help build intuition regarding channel parameter tradeoffs, it is advantageous to obtain a rough idea of how the contours of constant relative entropy $\mathcal{D}(\,\rho\,\|\,\phi\,)$ behave, for fixed ϕ, as ρ is varied. Furthermore, it will turn out that due to symmetries in the relative entropy, we frequently will only need to understand the relative entropy behavior in a plane of the Bloch sphere, which we choose to be the Bloch X-Y plane. In Figure 2.2, we plot a few contour lines for $\mathcal{D}(\,\rho\,\|\,\phi\;=\;\frac{1}{2}\,\mathcal{I}_2\,)$ in the X-Y Bloch sphere plane. In the figures that follow, we shall mark the location of ϕ with an asterisk (*). The contour values for $\mathcal{D}(\,\rho\,\|\,\phi\,)$ are shown in the plot title. The smallest value of $\mathcal{D}(\,\rho\,\|\,\phi\,)$ corresponds to the contour closest to the location of ϕ. The largest value of $\mathcal{D}(\,\rho\,\|\,\phi\,)$ corresponds to the outermost contour. For $\phi \;=\; \frac{1}{2}\mathcal{I}_2$, the location of ϕ is the Bloch sphere origin.

As an example of how these contour lines change as ϕ moves away from the maximally mixed state $\phi \;=\; \frac{1}{2}\mathcal{I}_2$, or equivalently as q becomes non-zero, we give contour plots below for $q \;\neq\; 0$. We let $\phi \;=\; \frac{1}{2}\,\{\,\mathcal{I}_2 \;+\; q\,\sigma_y\,\}$ with corresponding Bloch vector $\vec{\mathcal{V}} \;=\; \begin{bmatrix} 0 \\ q \\ 0 \end{bmatrix}$. The asterisk (*) in these plots denotes the location of $\vec{\mathcal{V}}$. The dashed outer contour is a radius equal to one, indicating where the pure states lie.

The two dimensional plots of $\mathcal{D}(\,\rho\,\|\,\phi\,)$ shown above tell us about the *three dimensional* nature of $\mathcal{D}(\,\rho\,\|\,\phi\,)$. To see why, first note that we can always rotate the Bloch sphere X-Y-Z axes to arrange for $\phi \;\equiv\; \vec{\mathcal{V}} \;\rightarrow\; \vec{q}$ to lie on the Y axis, as the density matrices ϕ are

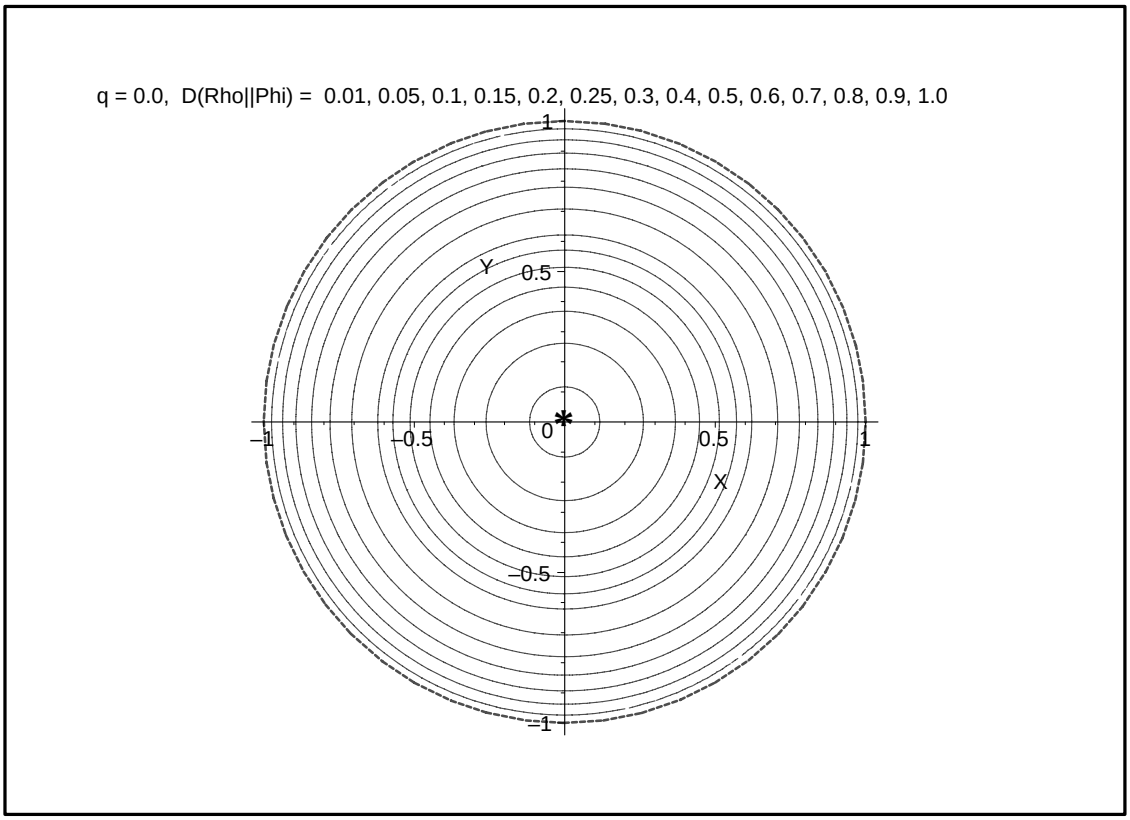

Figure 2.2: Contours of constant relative entropy $\mathcal{D}(\rho\|\phi)$ as a function of ρ in the Bloch sphere X-Y plane for the fixed density matrix $\phi = \frac{1}{2}\mathcal{I}_2$.

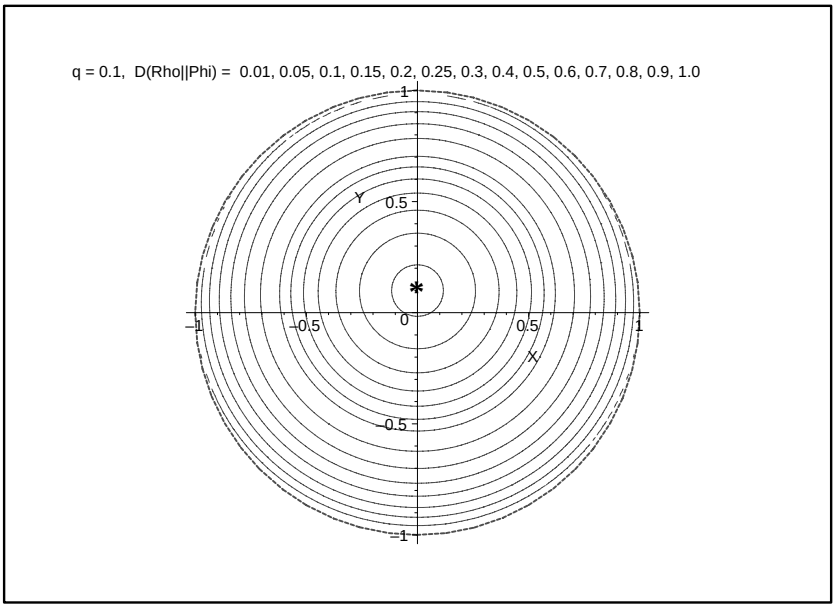

Figure 2.3: Contours of constant relative entropy $\mathcal{D}(\rho\|\phi)$ as a function of ρ in the Bloch sphere X-Y plane for the fixed density matrix $\phi = \frac{1}{2}\{\mathcal{I}_2 + 0.1\,\sigma_y\}$.

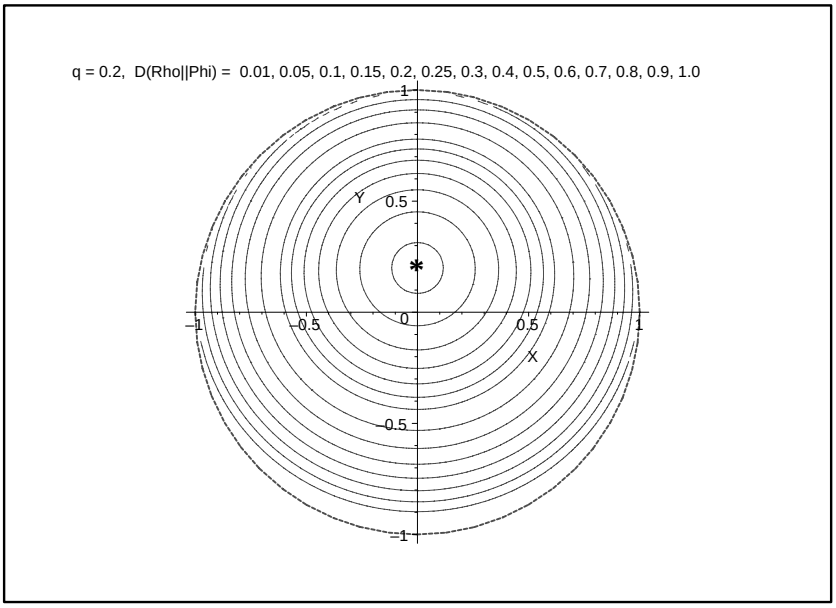

Figure 2.4: Contours of constant relative entropy $\mathcal{D}(\rho\|\phi)$ as a function of ρ in the Bloch sphere X-Y plane for the fixed density matrix $\phi = \frac{1}{2}\{\mathcal{I}_2 + 0.2\,\sigma_y\}$.

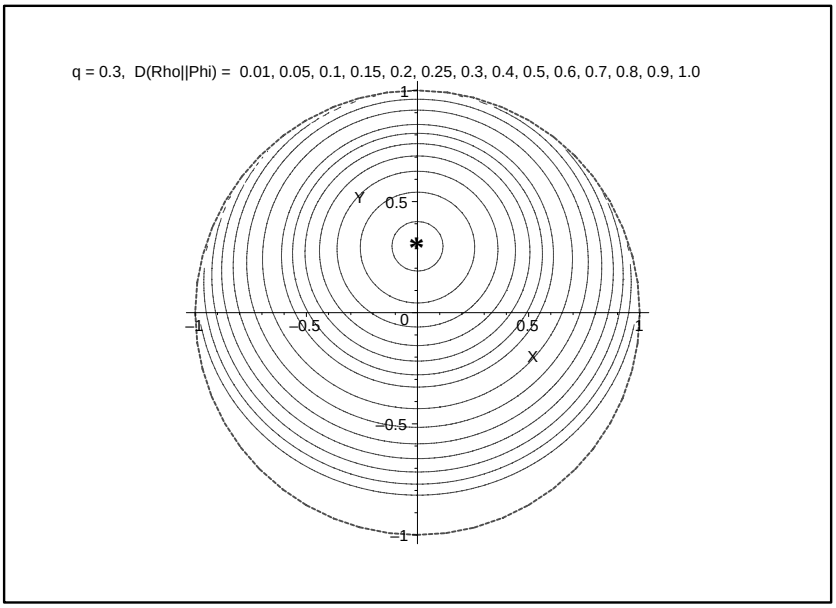

Figure 2.5: Contours of constant relative entropy $\mathcal{D}(\rho\|\phi)$ as a function of ρ in the Bloch sphere X-Y plane for the fixed density matrix $\phi = \frac{1}{2}\{\mathcal{I}_2 + 0.3\,\sigma_y\}$.

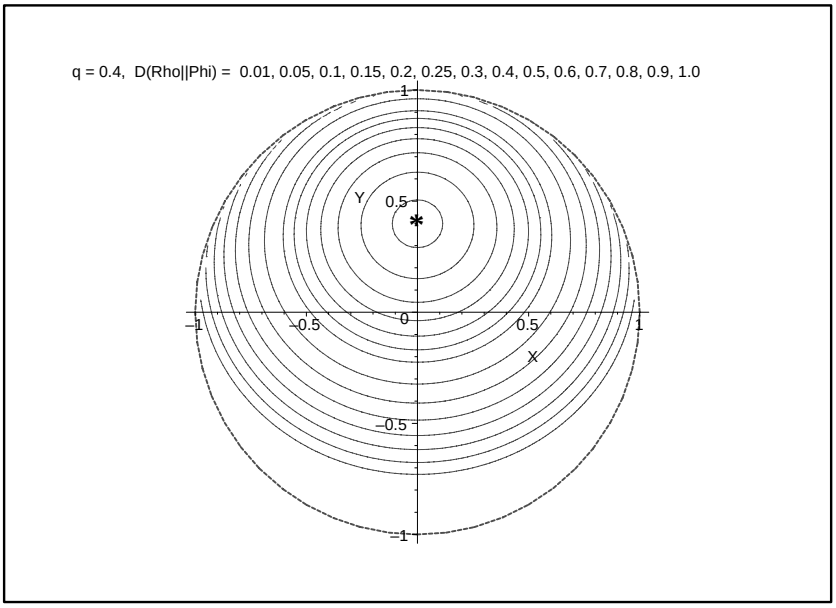

Figure 2.6: Contours of constant relative entropy $\mathcal{D}(\rho\|\phi)$ as a function of ρ in the Bloch sphere X-Y plane for the fixed density matrix $\phi = \frac{1}{2}\{\mathcal{I}_2 + 0.4\,\sigma_y\}$.

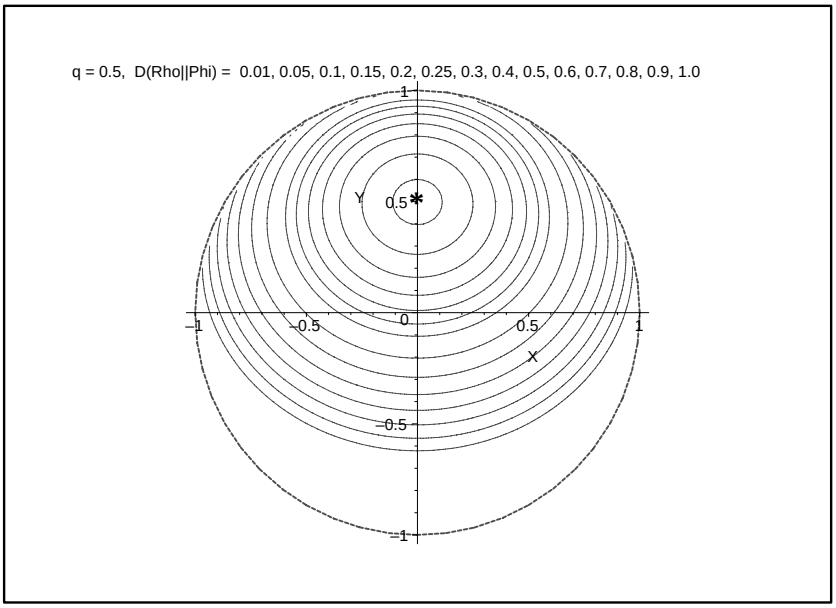

Figure 2.7: Contours of constant relative entropy $\mathcal{D}(\rho\|\phi)$ as a function of ρ in the Bloch sphere X-Y plane for the fixed density matrix $\phi = \frac{1}{2}\{\mathcal{I}_2 + 0.5\,\sigma_y\}$.

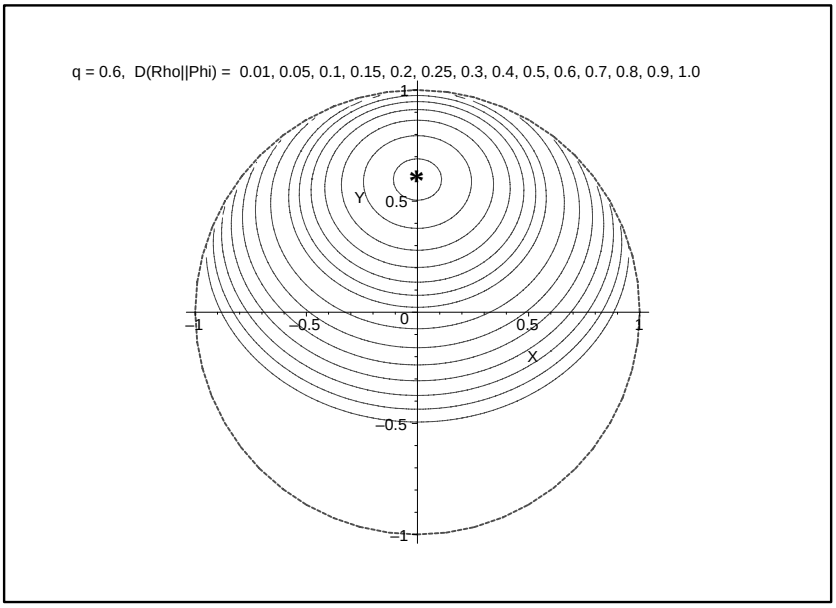

Figure 2.8: Contours of constant relative entropy $\mathcal{D}(\rho\|\phi)$ as a function of ρ in the Bloch sphere X-Y plane for the fixed density matrix $\phi = \frac{1}{2}\{\mathcal{I}_2 + 0.6\,\sigma_y\}$.

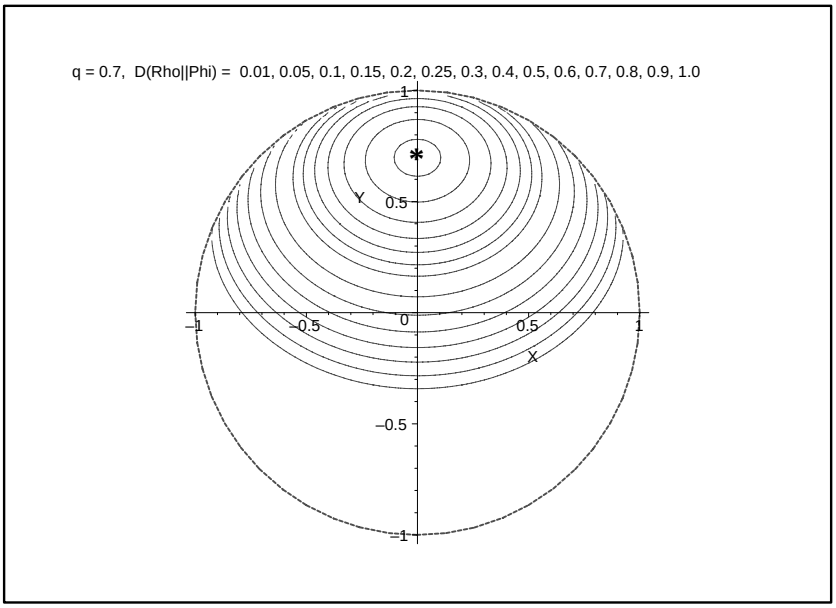

Figure 2.9: Contours of constant relative entropy $\mathcal{D}(\rho\|\phi)$ as a function of ρ in the Bloch sphere X-Y plane for the fixed density matrix $\phi = \frac{1}{2}\{\mathcal{I}_2 + 0.7\,\sigma_y\}$.

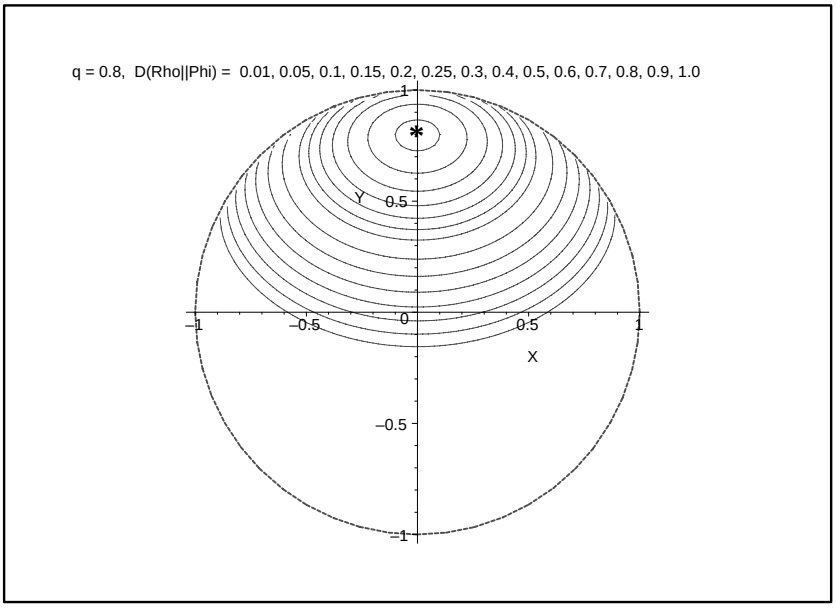

Figure 2.10: Contours of constant relative entropy $\mathcal{D}(\rho\|\phi)$ as a function of ρ in the Bloch sphere X-Y plane for the fixed density matrix $\phi = \frac{1}{2}\{\mathcal{I}_2 + 0.8\,\sigma_y\}$.

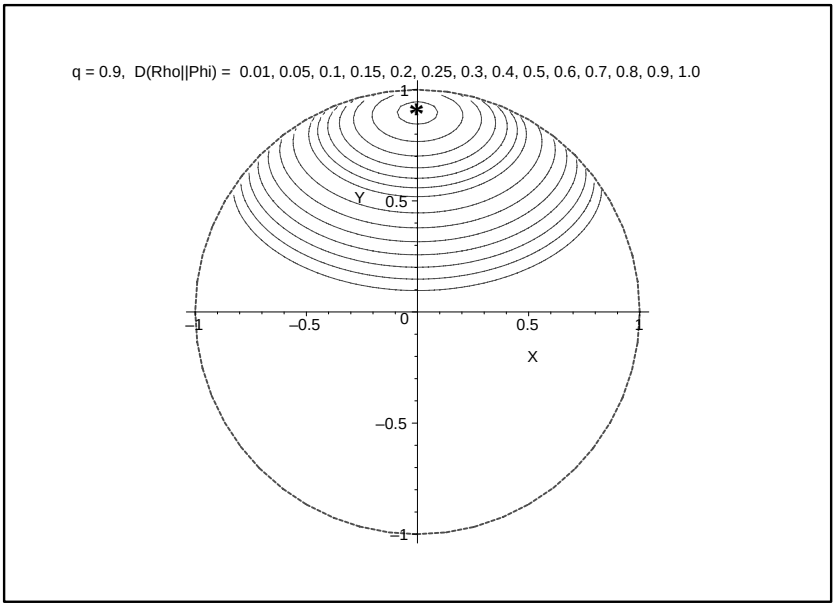

Figure 2.11: Contours of constant relative entropy $\mathcal{D}(\rho\|\phi)$ as a function of ρ in the Bloch sphere X-Y plane for the fixed density matrix $\phi = \frac{1}{2}\{\mathcal{I}_2 + 0.9\,\sigma_y\}$.

shown in Figures 2.2 through 2.11 above.

Second, recall that our description of $\mathcal{D}(\rho\,\|\,\phi)$ is a function of the three variables $\{r, q, \theta\}$ only, which were defined above as the length of the Bloch vectors corresponding to the density matrices ρ and ϕ respectively, and the angle between these Bloch vectors.

$$\mathcal{D}(\rho\,\|\,\phi) \equiv f(r, q, \theta).$$

This means the two dimensional curves of constant $\mathcal{D}(\rho\,\|\,\phi)$ can be rotated about the Y axis as surfaces of revolution, to yield three dimensional surfaces of constant $\mathcal{D}(\rho\,\|\,\phi)$. (In these two and three dimensional plots, the first argument of $\mathcal{D}(\rho\,\|\,\phi)$, ρ, is being varied, while the second argument, ϕ, is being held fixed at a point on the Y axis.)

Our two dimensional plots above give us a good idea of the three dimensional behavior of the surfaces of constant relative entropy about the density matrix ϕ occupying the second slot in $\mathcal{D}(\cdots\|\cdots)$. A picture emerges of slightly warped "eggshells" nested like Russian

dolls inside each other, *roughly* centered on ϕ. A mental picture of the behavior of $\mathcal{D}(\,\rho\,\|\,\phi\,)$ is useful because in what follows we shall superimpose the KRSW channel ellipsoid(s) onto Figures 2.2 through 2.11 above. [4] By moving $\vec{\mathcal{V}}$ (the asterisk) around in these pictures, we shall adjust the contours of constant $\mathcal{D}(\,\rho\,\|\,\phi\,)$, and thereby *graphically* determine the HSW channel capacity, optimum (output) signalling states, and corresponding a priori signalling probabilities. The resulting intuition we gain from these pictures will help us understand channel parameter tradeoffs.

V Linear Channels

Recall the KRSW specification of a qubit channel in terms of the six real parameters $\{\ t_x\,,\,t_y\,,\,t_z\,,\,\lambda_x\,,\,\lambda_y\,,\,\lambda_z\ \}$ as defined on page 24 in Chapter 1. A linear channel is one where $\lambda_x\ =\ \lambda_y\ =\ 0$, but $\lambda_z\ \neq\ 0$. The shift quantities t_k can be any real number, up to the limits imposed by the requirement that the map be completely positive. For more details on the complete positivity requirements of qubit maps, please see [26].

A linear channel is a simple system that illustrates the basic ideas behind our graphical approach to determining the HSW channel capacity $\mathcal{C}_1$. Recall the relative entropy formulation for $\mathcal{C}_1$.

$$\mathcal{C}_1\ =\ \mathrm{Max}_{[all\ possible\ \{p_k\,,\,\varphi_k\}]}\quad \sum_k\ p_k\ \mathcal{D}\,(\,\mathcal{E}(\,\varphi_k\,)\,||\,\mathcal{E}(\,\varphi\,)\,)\,,$$

where the φ_k are the quantum states input to the channel and $\varphi\ =\ \sum_k\ p_k\,\varphi_k$. We call an ensemble of states $\{\,p_k\,,\,\varrho_k\ =\ \mathcal{E}(\varphi_k)\,\}$ an optimal ensemble if this ensemble achieves $\mathcal{C}_1$.

As discussed on page 19, Schumacher and Westmoreland showed the above maximization to determine $\mathcal{C}_1$ is equivalent to the following min-max criterion:

$$\mathcal{C}_1\ =\ \mathrm{Min}_{\,\phi}\qquad \mathrm{Max}_{\,\varrho_k}\qquad \mathcal{D}\,(\,\varrho_k\,\|\,\phi\,)\,,$$

where ϱ_k is a density matrix on the surface of the channel ellipsoid, and ϕ is a density

[4] See Section IX.a of Chapter 1 for the KRSW qubit channel formalism.

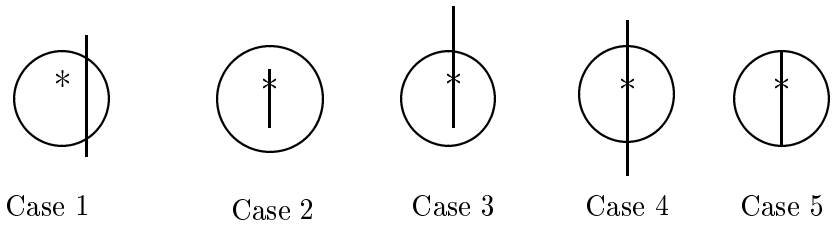

Figure 2.12: Scenarios for the intersection of the optimum relative entropy contour with a linear channel ellipsoid.

matrix in the convex hull of the channel ellipsoid. For the linear channel, the channel ellipsoid is a line segment of length 2 λ_z centered on $\{ t_x , t_y , t_z \}$. Thus, both ϱ_k and ϕ must lie somewhere along this line segment. Furthermore, Schumacher and Westmoreland tell us that ϕ must be expressible as a convex combination of the ϱ_k which satisfy the above min-max[16].

To graphically implement the min-max criterion, we overlay the channel ellipsoid on the contour plots of relative entropy previously found. We wish to determine the location of the optimum ϕ and the optimum relative entropy contour that achieves the min-max. The generic overlap scenarios are shown above, labeled Cases 1 - 5. From our plots of relative entropy, we know that contours of relative entropy are *roughly* circular about ϕ. We denote the location of ϕ above by an asterisk (*). The permissible ϱ_k are those density matrices at the intersection of the relative entropy contour and the channel ellipsoid, here a line segment.

Let us examine the five cases shown above, seeking the optimum ϕ and the relative entropy contour corresponding to $\mathcal{C}_1$ (the circles below), by eliminating those cases which do not make sense in light of the minimization-maximization above.

Case 1 is not an acceptable configuration because ϕ does not lie inside the channel ellipsoid, meaning for the linear channel, ϕ does not lie on the line segment. Case 2 is not acceptable because there are no permissible ϱ_k, since the relative entropy contour does not intersect the channel ellipsoid line segment anywhere. Case 3 is not acceptable because Schumacher and

Westmoreland tell us that ϕ must be expressible as a convex combination of the ϱ_k density matrices which satisfy the above min-max requirement. There is only one permissible ϱ_k density matrix in Case 3, and since, as seen in the diagram for Case 3, $\phi \neq \varrho_1$, we do not have an acceptable configuration. Case 4 at first appears acceptable. However, here we do not achieve the maximization in the min-max relation, since we can do better by using a relative entropy contour with a larger radii. Case 5 is the ideal situation. The relative entropy contour intersects both of the line segment endpoints. Taking a larger radius relative entropy contour does not give us permissible ϱ_k, since we would obtain Case 2 with a larger radii. For Case 5, if we moved ϕ as we increased the relative entropy contour, we would obtain Case 3, again an unacceptable configuration. In Case 5, using the two ϱ_k that lie at the intersection of the relative entropy contour and the channel ellipsoid line segment, we can form a convex combination of these ϱ_k that equals ϕ. Case 5 is the best we can do, meaning Case 5 yields the largest radius relative entropy contour which satisfies the Schumacher-Westmoreland requirements. The value of this largest radii relative entropy contour is the HSW channel capacity we seek, $\mathcal{C}_1$.

We now restate Case 5 in Bloch vector notation. We shall associate the Bloch vector $\vec{\mathcal{V}}$ with ϕ, and the Bloch vectors $\vec{\mathcal{W}}_k$ with the ϱ_k density matrices. For the linear channel, from our analysis above which resulted in Case 5, we know that $\vec{\mathcal{V}}$ must lie on the line segment between the two endpoint vectors $\vec{\mathcal{W}}_+$ and $\vec{\mathcal{W}}_-$. (Note that from here on, we shall drop the tilde $\tilde{}$ we were previously using to denote channel output Bloch vectors, as almost all the Bloch vectors we shall talk about below are channel output Bloch vectors. The few instances when this is not the case shall be obvious.)

For a general linear channel, the KRSW ellipsoid channel parameters satisfy

$$\{\, t_x \neq 0\,,\; t_y \neq 0\,,\; t_z \neq 0\,,\; \lambda_x = 0\,,\; \lambda_y = 0\,,\; \lambda_z \neq 0\,\}.$$

Thus, we can explicitly determine the Bloch vectors $\vec{\mathcal{W}}_+$ and $\vec{\mathcal{W}}_-$, which we write below.

$$\rho_+ \rightarrow \vec{\mathcal{W}}_+ = \begin{bmatrix} t_x \\ t_y \\ t_z + \lambda_z \end{bmatrix}, \quad and \quad \rho_- \rightarrow \vec{\mathcal{W}}_- = \begin{bmatrix} t_x \\ t_y \\ t_z - \lambda_z \end{bmatrix}.$$

Note that the t_k and λ_z are real numbers, and any of them may be negative.

The Bloch vector $\vec{\mathcal{V}}$ however requires more work. We parameterize the Bloch sphere vector $\vec{\mathcal{V}}$ corresponding to ϕ by the real number α, specifying a position for $\vec{\mathcal{V}}$ along the line segment between $\vec{\mathcal{W}}_+$ and $\vec{\mathcal{W}}_-$.

$$\phi \rightarrow \vec{\mathcal{V}} = \begin{bmatrix} t_x \\ t_y \\ t_z + \alpha\,\lambda_z \end{bmatrix}.$$

Here $\alpha \in [-1, 1]$. Now recall that the Schumacher-Westmoreland maximal distance property (see property # I in Section VI.a) tells us that $D(\rho_+||\phi) = D(\rho_-||\phi)$. To find $\vec{\mathcal{V}}$, we shall apply the formula we have derived for relative entropy in the Bloch representation to $D(\rho_+||\phi) = D(\rho_-||\phi)$, and solve for α. The details are in Appendix B.

V.a A Simple Linear Channel Example

To illustrate the ideas presented above, we take as a simple example the linear channel with channel parameters: $\{\ t_x = 0\ ,\ t_y = 0\ ,\ t_z = 0.2\ ,\ \lambda_x = 0\ ,\ \lambda_y = 0\ ,\ \lambda_z = 0.4\ \}$. Because the channel is linear with $t_x = t_y = 0$, we shall be able to easily solve for $\vec{\mathcal{V}}$ and $\mathcal{C}_1$.

We define the real numbers r_+ and r_- as the Euclidean distance in the Bloch sphere from the Bloch sphere origin to the Bloch vectors $\vec{\mathcal{W}}_+$ and $\vec{\mathcal{W}}_-$. That is, r_+ and r_- are the magnitudes of the Bloch vectors $\vec{\mathcal{W}}_+$ and $\vec{\mathcal{W}}_-$ defined above. For the channel parameter numbers given, we find $r_+ = \|\ 0.4 + 0.2\ \| = 0.6$ and $r_- = \|\ 0.2 - 0.4\ \| = 0.2$. We similarly define q to be the magnitude of the Bloch vector $\vec{\mathcal{V}}$.

To find $\vec{\mathcal{V}}$, we shall apply the formula we have derived for relative entropy in the Bloch representation to $D(\rho_+||\phi) = D(\rho_-||\phi)$, or in Bloch sphere notation, $D(\vec{\mathcal{W}}_+ \,||\, \vec{\mathcal{V}}) = D(\vec{\mathcal{W}}_- \,||\, \vec{\mathcal{V}})$. The formula for relative entropy derived in Appendix A is $\mathcal{D}(\,\varrho_k \,\|\, \phi\,) =$

$$\frac{1}{2}\log_2\left(1 - r_k^2\right) + \frac{r_k}{2}\log_2\left(\frac{1+r_k}{1-r_k}\right) - \frac{1}{2}\log_2\left(1-q^2\right) - \frac{\vec{\mathcal{W}}_k \bullet \vec{\mathcal{V}}}{2\,q}\log_2\left(\frac{1+q}{1-q}\right)$$

$$= \frac{1}{2}\log_2\left(1 - r_k^2\right) + \frac{r_k}{2}\log_2\left(\frac{1+r_k}{1-r_k}\right) - \frac{1}{2}\log_2\left(1-q^2\right) - \frac{r_k\,\cos(\theta_k)}{2}\log_2\left(\frac{1+q}{1-q}\right).$$

where θ_k is the angle between $\vec{\mathcal{W}}_k$ and $\vec{\mathcal{V}}$. Intuitively, one notes that the nearly circular relative entropy contours about $\phi \equiv \vec{\mathcal{V}}$ tells us that $\vec{\mathcal{V}} \approx \frac{\vec{\mathcal{W}}_+ + \vec{\mathcal{W}}_-}{2}$. Given the channel parameter numbers, this fact about $\vec{\mathcal{V}}$, together with the linear nature of the channel ellipsoid, tell us that $\theta_+ = 0$ and $\theta_- = \pi$, so that $\cos(\,\theta_+\,) = 1$ and $\cos(\,\theta_-\,) = -1$. Using this information about the θ_k, and the identity

$$\tanh^{(-1)}[\,x\,] = \frac{1}{2}\log\left(\frac{1+x}{1-x}\right),$$

the relative entropy equality relation between the two endpoints of the linear channel can be solved for q.

$$q_{optimum} = \tanh\left[\frac{\frac{1}{2}\,\ln\left[\frac{1-r_+^2}{1-r_-^2}\right] + r_+\,\tanh^{(-1)}[r_+] - r_-\,\tanh^{(-1)}[r_-]}{r_+ + r_-}\right] = 0.2125.$$

Thus,

$$\vec{\mathcal{W}}_+ = \begin{bmatrix} 0 \\ 0 \\ 0.6 \end{bmatrix}, \quad \vec{\mathcal{W}}_- = \begin{bmatrix} 0 \\ 0 \\ -0.2 \end{bmatrix}, \quad and \quad \vec{\mathcal{V}} = \begin{bmatrix} 0 \\ 0 \\ 0.2125 \end{bmatrix}.$$

The corresponding density matrices are

$$\rho_+ = \frac{1}{2}\,(\,\mathcal{I}_2 + \vec{\mathcal{W}}_+ \bullet \vec{\sigma}\,), \quad \rho_- = \frac{1}{2}\,(\,\mathcal{I}_2 + \vec{\mathcal{W}}_- \bullet \vec{\sigma}\,), \quad \phi = \frac{1}{2}\,(\,\mathcal{I}_2 + \vec{\mathcal{V}} \bullet \vec{\sigma}\,).$$

This yields $\mathcal{D}(\,\rho_+ \,\|\, \phi\,) = \mathcal{D}(\,\rho_- \,\|\, \phi\,) = 0.1246$. Thus, the HSW channel capacity $\mathcal{C}_1$ is 0.1246. The location of the two density matrices ρ_+ and ρ_- are shown in Figure 2.13 below

as **O**.

Furthermore, the Schumacher-Westmoreland analysis tells us that the two states ρ_+ and ρ_- must average to ϕ, in the sense that if p_+ and p_- are the a priori probabilities of the two output signal states, then $p_+ \, \rho_+ \, + \, p_- \, \rho_- \, = \, \phi$. In our Bloch sphere notation, this relationship becomes $p_+ \, \vec{\mathcal{W}}_+ \, + \, p_- \, \vec{\mathcal{W}}_- \, = \, \vec{\mathcal{V}}$. The asterisk (*****) in Figure 2.13 below shows the position of $\vec{\mathcal{V}}$.

Another relation relating the a priori probabilities p_+ and p_- is $p_+ \, + \, p_- \, = \, 1$. Using these two equations, we can solve for the a priori probabilities p_+ and p_-. For our example,

$$p_+ \vec{\mathcal{W}}_+ \, + \, p_- \vec{\mathcal{W}}_- \, = \, p_+ \begin{bmatrix} 0 \\ 0 \\ 0.6 \end{bmatrix} \, + \, p_- \begin{bmatrix} 0 \\ 0 \\ -0.2 \end{bmatrix} \, = \, \vec{\mathcal{V}} \, = \, \begin{bmatrix} 0 \\ 0 \\ 0.2125 \end{bmatrix}.$$

Solving for p_+ and p_- yields $p_+ \, = \, 0.5156$ and $p_- \, = \, 0.4844$.

Note that here we have found the optimum *output* signal states ρ_+ and ρ_-. From these one can find the optimum *input* signal states by finding the states φ_+ and φ_- which map to the respective optimum output states ρ_+ and ρ_-. In our example above, these are $\varphi_+ \, \rightarrow \, \vec{\mathcal{W}}_+^{Input} \, = \, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ and $\varphi_- \, \rightarrow \, \vec{\mathcal{W}}_-^{Input} \, = \, \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}$.

For the general linear channel, where any or all of the t_k can be non-zero, we can reduce the capacity calculation to the solution of a single, one dimensional transcendental equation. (Please see Appendix B for the full derivation.)

Define

$$r_+^2 \, = \, t_x^2 \, + \, t_y^2 \, + \, (\, t_z \, + \, \lambda_z \,)^2.$$

$$q^2 \, = \, t_x^2 \, + \, t_y^2 \, + \, (\, t_z \, + \, \beta \, \lambda_z \,)^2.$$

$$r_-^2 \, = \, t_x^2 \, + \, t_y^2 \, + \, (\, t_z \, - \, \lambda_z \,)^2.$$

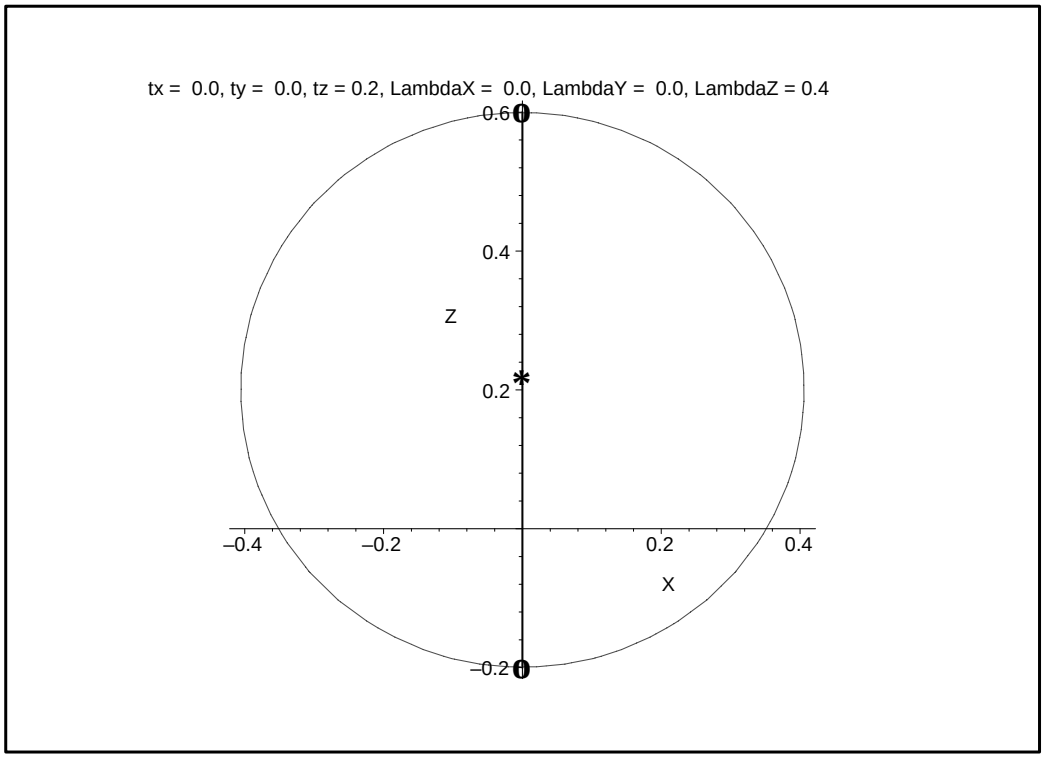

Figure 2.13: The intersection in the Bloch sphere X-Z plane of a linear channel ellipsoid and the optimum relative entropy contour. The optimum output signal states are shown as **O**.

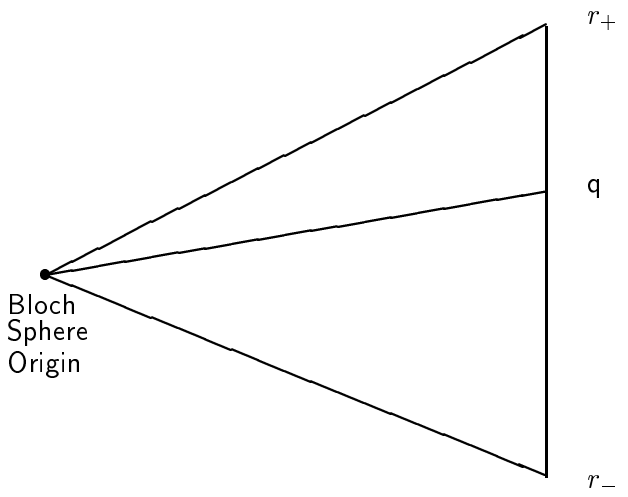

Figure 2.14: Definition of the Bloch vectors $\vec{r}_+$, $\vec{q}$, and $\vec{r}_-$ used in the derivation below.

The two quantities r_+ and r_- are the Euclidean distances from the Bloch sphere origin to the signalling states $\rho_+ \equiv \vec{\mathcal{W}}_+$ and $\rho_- \equiv \vec{\mathcal{W}}_-$, respectively. The quantity q is the Euclidean distance from the Bloch sphere origin to the density matrix $\phi \equiv \vec{\mathcal{V}}$. We define the three Bloch vectors $\vec{r}_+$, $\vec{q}$ and $\vec{r}_-$ in Figure 2.14, and refer to their respective magnitudes as r_+, q, and r_-.

We solve the transcendental equation below for β.

$$\frac{4\,\lambda_z\,(t_z\,+\,\beta\lambda_z\,)\,\tanh^{(-1)}(\,q\,)}{q}\;=$$

$$2\,r_+\,\tanh^{(-1)}(r_+)\;-\;2\,r_-\,\tanh^{(-1)}(r_-)\;+\;\ln(\,1\,-\,r_+^2\,)\;-\;\ln(\,1\,-\,r_-^2\,).$$

Note that q is a function of β, while r_+ and r_- are not. Thus, the right hand side remains constant while β is varied. The smooth nature of the functions of β on the left hand side allow a solution for β to be found fairly easily.

As in our simpler linear channel example above, we have

$$\vec{\mathcal{W}}_+ \;=\; \begin{bmatrix} t_x \\ t_y \\ t_z\;+\;\lambda_z \end{bmatrix}, \quad \vec{\mathcal{W}}_- \;=\; \begin{bmatrix} t_x \\ t_y \\ t_z\;-\;\lambda_z \end{bmatrix}, \quad \textit{and} \quad \vec{\mathcal{V}} \;=\; \begin{bmatrix} t_x \\ t_y \\ t_z + \beta\,\lambda_z \end{bmatrix}$$

where $\beta \in (-1, 1)$. The corresponding density matrices are:

$$\rho_+ = \frac{1}{2}(\mathcal{I}_2 + \vec{\mathcal{W}}_+ \bullet \vec{\sigma}), \quad \rho_- = \frac{1}{2}(\mathcal{I}_2 + \vec{\mathcal{W}}_- \bullet \vec{\sigma}), \quad \phi = \frac{1}{2}(\mathcal{I}_2 + \vec{\mathcal{V}} \bullet \vec{\sigma}).$$

The channel capacity $\mathcal{C}_1$ is found from the relations

$$D(\rho_+||\phi) = D(\rho_-||\phi) = \chi_{optimum} = \mathcal{C}_1 .$$

The a priori signaling probabilities are found by solving the simultaneous probability equations $p_+ + p_- = 1$, and

$$p_+\vec{\mathcal{W}}_+ + p_-\vec{\mathcal{W}}_- = p_+ \begin{bmatrix} t_x \\ t_y \\ t_z + \lambda_z \end{bmatrix} + p_- \begin{bmatrix} t_x \\ t_y \\ t_z - \lambda_z \end{bmatrix} = \vec{\mathcal{V}} = \begin{bmatrix} t_x \\ t_y \\ t_z + \beta\,\lambda_z \end{bmatrix}.$$

This leads to a second probability equation of $p_+ - p_- = \beta$, yielding:

$$p_+ = \frac{1 + \beta}{2} \qquad \textit{and} \qquad p_- = \frac{1 - \beta}{2}.$$

V.b A More General Linear Channel Example

In the simple linear channel example above, we used $\{ t_x = t_y = 0 , \lambda_x = \lambda_y = 0 \}$. This choice yielded a rotational symmetry about the Z - axis which assured us the location of the optimum average output density matrix $\rho = p_+\rho_+ + p_-\rho_-$ was on the Z - axis. We used this fact to advantage in predicting the angles $\theta_{\{+,-\}}$, where $\theta_{\{+,-\}}$ was the angle between $\vec{\mathcal{W}}_{\{+,-\}}$ and $\vec{\mathcal{V}}$. Since we knew $\vec{\mathcal{W}}_{\{+,-\}}$ lay on the Z - axis, we found $\theta_+ = 0$ and $\theta_- = \pi$, simplifying the $\cos\left(\theta_{\{+,-\}}\right)$ terms in the relative entropy expressions for $D(\rho_+||\phi)$ and $D(\rho_-||\phi)$. In general, we do not have values of ± 1 for $\cos\left(\theta_{\{+,-\}}\right)$, and this complicates finding a solution for the linear channel relation $D(\rho_+||\phi) = D(\rho_-||\phi)$.

A more general linear channel example is one where the parameters $\{ t_x , t_y , t_z \}$ are all non-zero. Consider the parameter set $\{ t_x = 0.1, t_y = 0.2, t_z = 0.3, \lambda_x =$

$0,\ \lambda_y = 0,\ \lambda_z = 0.4\ \}$. Solving the transcendental equation derived in Appendix B yields $\beta = 0.0534$ and $\vec{\mathcal{V}} = \begin{bmatrix} 0.1 \\ 0.2 \\ 0.3214 \end{bmatrix}$. Using the density matrix ϕ calculated from the Bloch vector $\vec{\mathcal{V}}$ gives us a HSW channel capacity $\mathcal{C}_1$ of $D(\rho_+||\phi) = D(\rho_-||\phi) = 0.1365$.

As discussed above, $p_+ + p_- = 1$, and $p_+ - p_- = \beta$. Solving for p_+ and p_- yields $p_+ = 0.5267$ and $p_- = 0.4733$.

The optimum *input* Bloch vectors are:

$$\varphi_+ \rightarrow \vec{\mathcal{W}}_+^{Input} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \qquad and \qquad \varphi_- \rightarrow \vec{\mathcal{W}}_-^{Input} = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} .$$

The optimum *output* Bloch vectors are:

$$\rho_+ = \mathcal{E}(\varphi_+) \rightarrow \vec{\mathcal{W}}_+^{Output} = \begin{bmatrix} 0.1 \\ 0.2 \\ 0.7 \end{bmatrix} \qquad and \qquad \rho_- = \mathcal{E}(\varphi_-) \rightarrow \vec{\mathcal{W}}_-^{Output} = \begin{bmatrix} 0.1 \\ 0.2 \\ -0.1 \end{bmatrix} .$$

Below we show in Figure 2.15 and Figure 2.16 the $\{x, z\}$ and $\{y, z\}$ slices of the *linear* channel ellipsoid. One can see that the relative entropy curve $D(\rho \,\|\, \phi) = \mathcal{C}_1 = 0.1365$ touches the ellipsoid at two locations in both cross sections. (The $\{x, y\}$ cross section is trivial.)

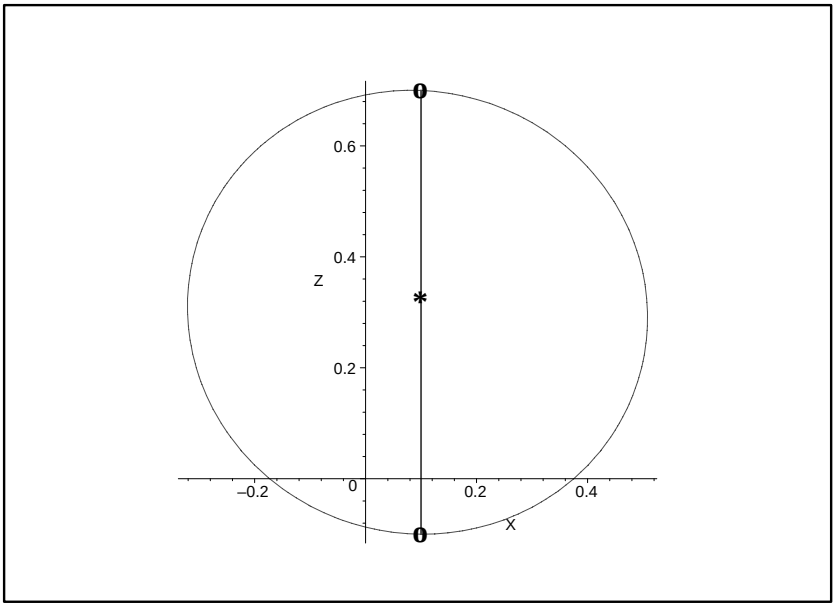

Figure 2.15: The intersection in the Bloch sphere X-Z plane of a linear channel ellipsoid and the optimum relative entropy contour. The optimum output signal states are shown as **O**.

VI Planar Channels

A planar channel is a quantum channel where two λ_k are non-zero, and one λ_k is zero. For a planar channel, the $\{\ t_k\ \}$ can have any values allowed by complete positivity. A planar channel restricts the possible output density matrices to lie in the plane in the Bloch sphere which is specified by the non-zero λ_k. In comparison to the linear channels discussed above, the planar channels additional output degree of freedom (planar has two non-zero λ_k versus a single linear non-zero λ_k) means a slightly different approach to determining $\mathcal{C}_1$ than that discussed for linear channels must be developed. As for linear channels, we seek to find the optimum density matrix $\phi \equiv \vec{\mathcal{V}}$ interior to the ellipsoid which minimizes the distance to the most "distant", in a relative entropy sense, point(s) on the ellipsoid surface. We shall find the optimum $\vec{\mathcal{V}}$ in two ways: graphically and iteratively. Both approaches utilize the following theorem from Schumacher and Westmoreland[16].

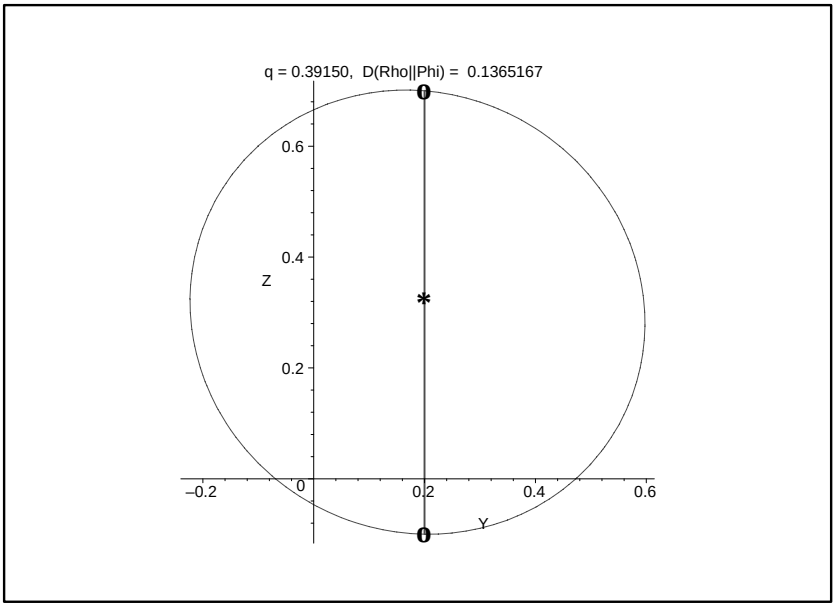

Figure 2.16: The intersection in the Bloch sphere Y-Z plane of a linear channel ellipsoid and the optimum relative entropy contour. The optimum output signal states are shown as **O**.

Theorem:

$$\mathcal{C}_1 \quad = \quad Min_{\phi} \quad Max_{\rho} \quad D(\,\rho \,\|\, \phi\,).$$

The maximum is taken over the *surface* of the ellipsoid, and the minimum is taken over the *interior* of the ellipsoid. In order to apply the min max formula above for $\mathcal{C}_1$ for planar channels, we need a result about the uniqueness of the average output ensemble density matrix $\rho \;=\; \sum_k \; p_k \, \rho_k$ for different optimal ensembles $\{\; p_k \;,\; \rho_k \;\}$.

Recall the following theorem proven in Section VIII.b of Chapter 1.

Theorem:

The density matrix ϕ which achieves the minimum in the min-max formula above for $\mathcal{C}_1$ is unique.

Below, we use the uniqueness of the average output ensemble density matrix in optimization

procedures which yield the HSW capacity of an arbitrary qubit channel.

VI.a Graphical Channel Optimization Procedure

We describe a graphical technique for finding $\phi_{optimum} \equiv \vec{\mathcal{V}}_{optimum}$. Recall the contour surfaces of constant relative entropy for various values of $\vec{\mathcal{V}}$ shown previously. We seek to adjust the location of $\vec{\mathcal{V}}$ inside the channel ellipsoid such that the largest possible contour value $\mathcal{D}_{max} = \mathcal{D}(\vec{\mathcal{W}} \| \vec{\mathcal{V}})$ touches the ellipsoid surface, and the remainder of the $\mathcal{D}_{max}$ contour surface lies entirely outside the channel ellipsoid. Our linear channel example illustrated this idea. In that example, the $\mathcal{D}_{max}$ contour intersects the "ellipsoid" at r_+ and r_-, and otherwise lies outside the line segment between r_+ and r_- representing the convex hull of $\mathcal{A}$. (Recall from the discussion of the Schumacher and Westmoreland paper in Section 2.2 that the points on the ellipsoid surface were defined as the set $\mathcal{A}$, and the interior of the ellipsoid, where $\vec{\mathcal{V}}$ lives, is the convex hull of $\mathcal{A}$.)

A good place to start is with $\vec{\mathcal{V}}_{initial} = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$. We then "tweak" $\vec{\mathcal{V}}$ as described above to find $\vec{\mathcal{V}}_{optimum}$. Note that $\vec{\mathcal{V}}_{optimum}$ should be near $\vec{\mathcal{V}}_{initial}$ because of the *almost* radial symmetry of $\mathcal{D}$ about $\vec{\mathcal{V}}$ as seen in Figures 2.2 through 2.11.

This technique is graphically implementing property IV in Section 2.2. In Bloch sphere notation, we have:

$$\mathcal{C}_1 = \mathrm{Min}_{\vec{\mathcal{V}}} \quad \mathrm{Max}_{\vec{\mathcal{W}}} \quad \mathcal{D}\left(\vec{\mathcal{W}} \| \vec{\mathcal{V}}\right),$$

where $\vec{\mathcal{W}}$ is on the channel ellipsoid surface and $\vec{\mathcal{V}}$ is in the interior of the ellipsoid. Moving $\vec{\mathcal{V}}$ from the optimum position described above will increase $Max_{\vec{\mathcal{W}}} \; \mathcal{D}\left(\vec{\mathcal{W}} \| \vec{\mathcal{V}}\right)$, since a larger contour value of $\mathcal{D}$ would then intersect the channel ellipsoid surface, thereby *increasing* $Max_{\vec{\mathcal{W}}} \; \mathcal{D}\left(\vec{\mathcal{W}} \| \vec{\mathcal{V}}\right)$. Yet $\vec{\mathcal{V}}$ should be adjusted to *minimize* $Max_{\vec{\mathcal{W}}} \; \mathcal{D}\left(\vec{\mathcal{W}} \| \vec{\mathcal{V}}\right)$.

VI.b Iterative Channel Optimization Procedure

For the iterative treatment, we outline an algorithm which converges to $\vec{\mathcal{V}}_{optimum}$. First, we need a lemma.

Lemma: Let $\vec{\mathcal{V}}$ and $\vec{\mathcal{W}}$ be any two Bloch sphere vectors. Define a third Bloch sphere vector $\vec{\mathcal{U}}$ as:

$$\vec{\mathcal{U}} \;=\; (\,1\;-\alpha\,)\,\vec{\mathcal{W}} \;+\; \alpha\,\vec{\mathcal{V}},$$

where $\alpha \;\in\; (0,1)$. Then

$$\mathcal{D}(\,\vec{\mathcal{W}}\,\|\,\vec{\mathcal{U}}\,) \;<\; \mathcal{D}(\,\vec{\mathcal{W}}\,\|\,\vec{\mathcal{V}}\,).$$

Proof:

By the joint convexity property of the relative entropy:[2]

$$\mathcal{D}(\,\{\,\alpha\,\rho_1 + (\,1\,-\,\alpha\,)\,\rho_2\,\}\,\|\,\{\,\alpha\,\phi_1 + (\,1\,-\,\alpha\,)\,\phi_2\,\}\,) \;\leq\; \alpha\,\mathcal{D}(\,\rho_1\,\|\,\phi_1\,) \;+\; (\,1\,-\,\alpha\,)\,\mathcal{D}(\,\rho_2\,\|\,\phi_2\,),$$

where $\alpha \;\in\; (0,1)$. Let $\rho_1 \;=\; \rho_2 \;\equiv\; \vec{\mathcal{W}}$, $\phi_1 \;\equiv\; \vec{\mathcal{V}}$ and $\phi_2 \;\equiv\; \vec{\mathcal{W}}$ with $\vec{\mathcal{U}} \;=\; (\,1\,-\alpha\,)\,\vec{\mathcal{W}} + \alpha\,\vec{\mathcal{V}}$. We obtain:

$$\mathcal{D}(\,\vec{\mathcal{W}}\,\|\vec{\mathcal{U}}\,) \;=\; \mathcal{D}(\,\vec{\mathcal{W}}\,\|\,\alpha\,\vec{\mathcal{V}} + (\,1\,-\,\alpha\,)\,\vec{\mathcal{W}}\,) \;\leq\; \alpha\,\mathcal{D}(\,\vec{\mathcal{W}}\,\|\,\vec{\mathcal{V}}\,) \;+\; (\,1\,-\,\alpha\,)\,\mathcal{D}(\,\vec{\mathcal{W}}\,\|\,\vec{\mathcal{W}}\,)\,.$$

But $\mathcal{D}(\,\vec{\mathcal{W}}\,\|\,\vec{\mathcal{W}}\,) \;=\; 0$, by Klein's inequality[2]. Thus,

$$\mathcal{D}(\,\vec{\mathcal{W}}\,\|\vec{\mathcal{U}}\,) \;\leq\; \alpha\,\mathcal{D}(\,\vec{\mathcal{W}}\,\|\,\vec{\mathcal{V}}\,) \;<\; \mathcal{D}(\,\vec{\mathcal{W}}\,\|\,\vec{\mathcal{V}}\,),$$

since $\alpha \;\in\; (0,1)$.

$\triangle$ - *End of Proof.*

We use the lemma above to guide us in iteratively adjusting $\vec{\mathcal{V}}$ to converge towards $\vec{\mathcal{V}}_{optimal}$. Consider $\mathcal{D}(\,\vec{\mathcal{W}}\,\|\,\vec{\mathcal{V}}\,)$, where $\vec{\mathcal{W}} \;\in\; \mathcal{A}$ and $\vec{\mathcal{V}} \;\in\; \mathcal{B} \;\equiv$ the convex hull of $\mathcal{A}$. We seek to find

$\mathcal{C}_1$ in an iterative fashion. We do this by holding $\vec{\mathcal{V}}$ fixed, and finding one of the $\vec{\mathcal{W}}' \in \mathcal{A}$ which maximizes $\mathcal{D}(\vec{\mathcal{W}} \,\|\, \vec{\mathcal{V}})$. From our lemma above, if we now move $\vec{\mathcal{V}}$ towards $\vec{\mathcal{W}}'$, we shall cause $\mathcal{D}_{max}(\vec{\mathcal{V}}) = Max_{\vec{\mathcal{W}}}\, \mathcal{D}(\vec{\mathcal{W}} \,\|\, \vec{\mathcal{V}})$ to decrease. We steadily decrease $\mathcal{D}_{max}(\vec{\mathcal{V}})$ in this manner until we reach a point where any movement of $\vec{\mathcal{V}}$ will increase $\mathcal{D}_{max}(\vec{\mathcal{V}})$. Our uniqueness theorem above tells us there is only one $\vec{\mathcal{V}}_{optimal}$. Our lemma above tells us we cannot become stuck in a local minima in moving towards $\vec{\mathcal{V}}_{optimal}$. Thus, when we reach the point where any movement of $\vec{\mathcal{V}}$ will increase $\mathcal{D}_{max}(\vec{\mathcal{V}})$, we are done and have found $\vec{\mathcal{V}}_{final} = \vec{\mathcal{V}}_{optimum}$.

To summarize, we find the optimum $\vec{\mathcal{V}}$ using the following algorithm.

1) Generate a random starting point $\vec{\mathcal{V}}_{initial}$ in the interior of the ellipsoid ($\in \mathcal{B}$). (In actuality, since the contour surfaces of constant relative entropy are *roughly* spherical about $\vec{\mathcal{V}}$, a good place to start is $\vec{\mathcal{V}}_{initial} = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$.)

2) Determine the set of points $\{ \vec{\mathcal{W}}' \}$ on the ellipsoid surface most distant, in a relative entropy sense, from our $\vec{\mathcal{V}}$. This maximal distance is $\mathcal{D}_{max}(\vec{\mathcal{V}})$ defined above as $\mathcal{D}_{max}(\vec{\mathcal{V}}) = Max_{\vec{\mathcal{W}}'}\, \mathcal{D}(\vec{\mathcal{W}}' \,\|\, \vec{\mathcal{V}})$.

3) Choose at random one Bloch sphere vector from our maximal set of points $\{ \vec{\mathcal{W}}' \}$. Call this selected point $\widehat{\vec{\mathcal{W}}'}$. In the 3 real dimensional Bloch sphere space, make a small step from $\vec{\mathcal{V}}$ towards the surface point vector, $\widehat{\vec{\mathcal{W}}'}$. That is, update $\vec{\mathcal{V}}$ as follows:

$$\vec{\mathcal{V}}_{new} = (1 - \epsilon)\, \vec{\mathcal{V}}_{old} + \epsilon\, \widehat{\vec{\mathcal{W}}'}.$$

4) Loop by going back to step 2) above, using our new, updated $\vec{\mathcal{V}}_{new}$, and continue to loop until $\mathcal{D}_{max}$ is no longer changing.

This algorithm converges to $\phi_{optimum} \equiv \vec{\mathcal{V}}_{optimum}$, because we steadily proceed downhill minimizing $Max_{\vec{\mathcal{W}}} \quad \mathcal{D}\left(\vec{\mathcal{W}} \,\|\, \vec{\mathcal{V}}\right)$, and our lemma above tells us we can never get stuck in a local minima.

VI.c Planar Channel Example

We demonstrate the iterative algorithm above with a planar channel example. Let $\{\ t_x = 0.3,\ t_y = 0.1,\ t_z = 0,\ \lambda_x = 0.4,\ \lambda_y = 0.5,\ \lambda_z = 0\ \}$. The iterative algorithm outlined above yields $\vec{\mathcal{V}} = \begin{bmatrix} 0.3209 \\ 0.1112 \\ 0 \end{bmatrix}$ and a HSW channel capacity $\mathcal{C}_1 = \mathcal{D}_{optimum} = 0.1994$. Shown in Figure 2.17 is a plot of the planar channel ellipsoid contour and the relative entropy contour. The planar output channel ellipsoid contour is the inner, dashed curve. The contour of constant relative entropy $\mathcal{D}(\,\rho\,\|\,\phi\,) = \mathcal{D}_{optimum}$ is centered at $\vec{\mathcal{V}}$, which is marked in Figure 17 with an asterisk *****. One can see that the $\mathcal{D}_{max}$ curve intersects the ellipsoid curve at two points, marked with **O**, and these two points are the optimum channel output signals ρ_i^{Output}.

The optimum input and output signalling states for this channel were determined as described in Appendix D and are:

$$P_1 = 0.4869, \qquad \vec{W}_1^{Input} = \begin{bmatrix} -0.0207 \\ -0.9998 \\ 0 \end{bmatrix}, \qquad \vec{W}_1^{Output} = \begin{bmatrix} 0.2917 \\ -0.3999 \\ 0 \end{bmatrix}.$$

$$P_2 = 0.5131, \qquad \vec{W}_2^{Input} = \begin{bmatrix} 0.1215 \\ 0.9926 \\ 0 \end{bmatrix}, \qquad \vec{W}_2^{Output} = \begin{bmatrix} 0.3486 \\ 0.5963 \\ 0 \end{bmatrix}.$$

These signal states yield an average channel output Bloch vector $\vec{\mathcal{V}}$ of

$$\vec{\mathcal{V}} = P_1 \cdot \vec{W}_1^{Output} + P_2 \cdot \vec{W}_2^{Output} = \begin{bmatrix} 0.3209 \\ 0.1113 \\ 0 \end{bmatrix}.$$

Figure 2.17 below shows the location of the channel ellipsoid (the inner dashed curve), the contour of constant relative entropy (the solid curve) for $\mathcal{D} = 0.1994$, the location of the two optimum input pure states ρ_i^{Input}, (the two **O** states on the circle of radius one), and

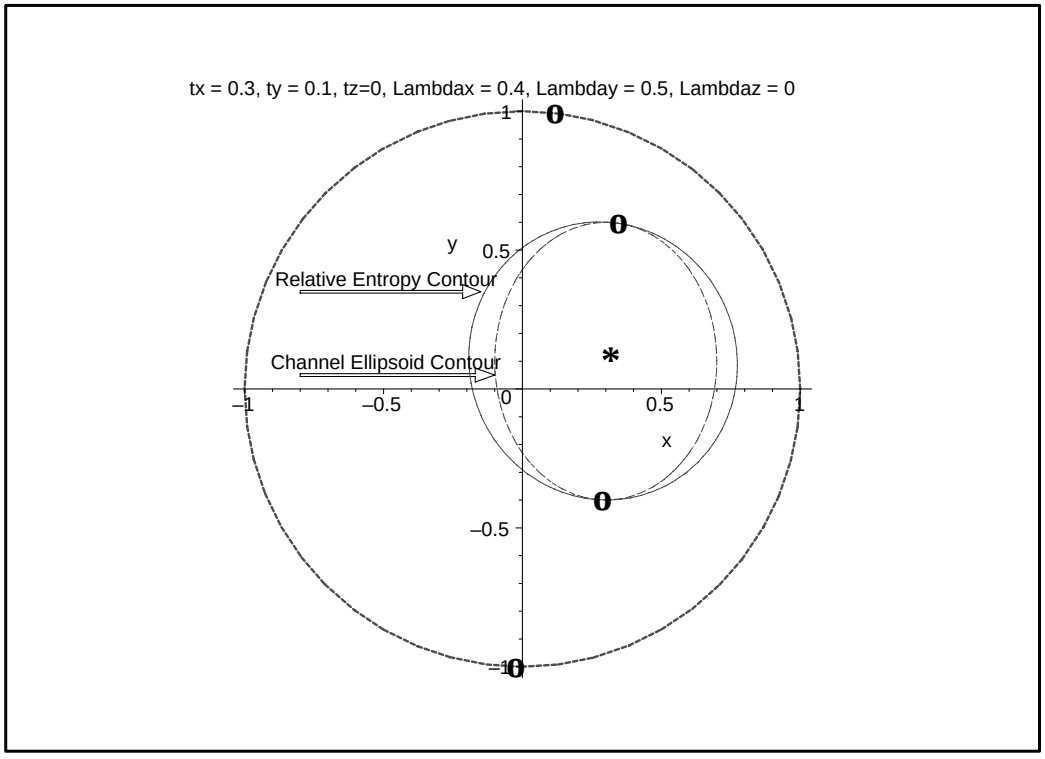

Figure 2.17: The intersection in the Bloch sphere X-Y plane of a planar channel ellipsoid (the inner dashed curve) and the optimum relative entropy contour (the solid curve). The two optimum input signal states (on the outer bold dashed Bloch sphere boundary curve) and the two optimum output signal states (on the channel ellipsoid *and* the optimum relative entropy contour curve) are shown as **O**.

the two optimum output signal states ρ_i^{Output}, also denoted by **O**, on the channel ellipsoid curve. Note that the optimum input signalling states are non-orthogonal.

Another useful picture is how the relative entropy changes as we make our way around the channel ellipsoid. We consider the Bloch X-Y plane in polar coordinates $\{ r , \theta \}$, where we measure the angle θ with respect to the origin of the Bloch X-Y plane axes. (Note that θ only fully ranges over $[0, 2\pi]$ when the origin of the Bloch sphere lies inside the channel ellipsoid.) The horizontal line at the top of the plot is the channel capacity $\mathcal{C}_1 = 0.1994$. Note that the two relative entropy peaks correspond to the locations of the two output optimum signalling states.

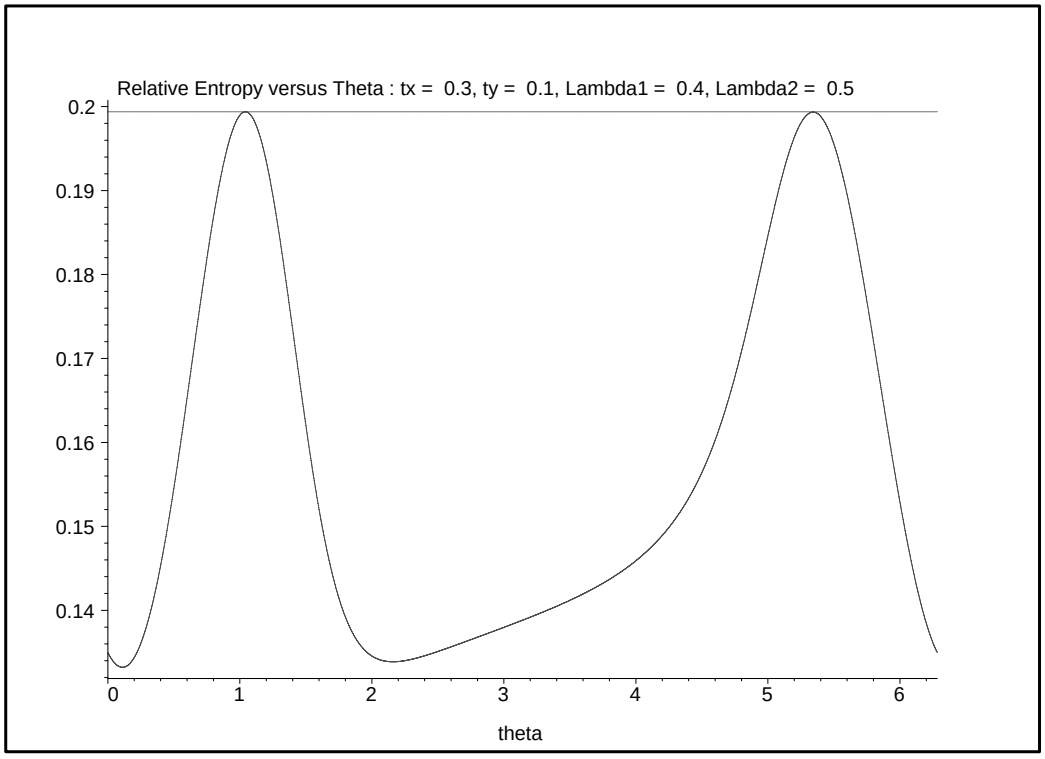

Figure 2.18: The change in $\mathcal{D}(\ \rho \parallel \phi \equiv * \ )$ as we move ρ around the channel ellipsoid. The angle theta is with respect to the Bloch sphere origin.

For this channel, the optimum channel capacity is achieved using an ensemble consisting of only two signalling states. Davies theorem tells us that for single qubit channels, an optimum ensemble need contain at most four signalling states. Using the notation of [40], we call C_2 the optimum output $\mathcal{C}_1$ HSW channel capacity attainable using only two input signalling states, C_3 is the optimum output $\mathcal{C}_1$ HSW channel capacity attainable using only three input signalling states, and C_4 is the optimum output $\mathcal{C}_1$ HSW channel capacity attainable using only four input signalling states. Thus, for this channel, we see that $C_2 = C_3 = C_4$. That is, for this channel, allowing more than two signalling states in your optimal ensemble does not yield additional channel capacity over an optimal ensemble with just two signalling states.

VII Unital Qubit Channels

Unital channels are quantum channels that map the identity to the identity: $\mathcal{E}(\mathcal{I}) = \mathcal{I}$. Due to this behavior, unital channels possess certain symmetries. In the ellipsoid picture, King and Ruskai [25] have shown that for unital channels, the $\{ t_k \}$ are zero. This yields an ellipsoid centered at the origin of the Bloch sphere. The resulting symmetry of such an ellipsoid will allow us to draw powerful conclusions.

First, recall that we know there exists at least one optimal signal ensemble, $\{ p_i , \rho_i \}$, which attains the HSW channel capacity $\mathcal{C}_1$. (See property III in Section VI.a.) Now consider the symmetry evident in the formula we have derived for the relative entropy for two single qubit density operators. We have:

$$\mathcal{D}(\rho \| \phi) = \mathcal{D}(\vec{\mathcal{W}} \| \vec{\mathcal{V}}) = f(r , q , \theta),$$

where $r = \| \vec{\mathcal{W}} \|$, $q = \| \vec{\mathcal{V}} \|$, and θ is the angle between $\vec{\mathcal{W}}$ and $\vec{\mathcal{V}}$. Thus, if $\rho_i \in \mathcal{A}$ and $\phi \in \mathcal{B}$, with $\mathcal{D}(\rho_i \| \sigma) = \mathcal{D}(\vec{\mathcal{W}}_i \| \vec{\mathcal{V}}) = \chi_{optimum} = \mathcal{C}_1$, then acting in $\mathcal{R}^3$, reflecting $\rho_i \equiv \vec{\mathcal{W}}_i$ and $\sigma \equiv \vec{\mathcal{V}}$ through the Bloch sphere origin to obtain $\rho_i' \equiv \vec{\mathcal{W}}_i'$ and $\sigma' \equiv \vec{\mathcal{V}}'$,

yields elements of $\mathcal{A}$ and $\mathcal{B}$ respectively. Furthermore, these transformed density matrices will also satisfy $\mathcal{D}(\,\rho_i'\,\|\,\sigma'\,) \;=\; \mathcal{D}(\,\vec{\mathcal{W}}_i'\,\|\,\vec{\mathcal{V}}'\,) \;=\; \chi_{optimum} \;=\; \mathcal{C}_1$, because r, q, and θ remain the same when we reflect through the Bloch sphere origin. That is, the symmetry of the unital channel ellipsoid about the Bloch sphere origin, corresponding to the density matrix $\frac{1}{2}\mathcal{I}_2$, together with the symmetry present in the qubit relative entropy formula yields a symmetry for the optimal signal ensemble $\{\,p_i\,,\,\rho_i\,\}$, where $\sigma \;=\; \sum_i\, p_i\,\rho_i$, or equivalently $\vec{\mathcal{V}} \;=\; \sum_i\, p_i\,\vec{\mathcal{W}}_i$. This symmetry indicates that for every optimal signal ensemble $\{\,p_i\,,\,\rho_i\,\}$, there exists another ensemble, $\{\,p_i'\,,\,\rho_i'\,\}$, obtained by reflection through the Bloch sphere origin. Since we know there exists at least one optimal signal ensemble, we must conclude that if $\sigma \;=\; \sum_i\, p_i\,\rho_i \;\neq\; \frac{1}{2}\mathcal{I}_2$, then two optimal ensembles exist with $\sigma \;\neq\; \sigma'$. However, by our uniqueness proof above, we are assured that $\sigma \;=\; \sum_i\, p_i\,\rho_i$ is a unique density matrix, regardless of the states $\{\,p_i\,,\,\rho_i\,\}$ used, as long as the states $\{\,p_i\,,\,\rho_i\,\}$ are an optimal ensemble. Thus we must conclude that $\sigma \;=\; \sum_i\, p_i\,\rho_i \;\equiv\; \frac{1}{2}\mathcal{I}_2$, since only the density matrix $\frac{1}{2}\mathcal{I}_2$ maps into itself upon reflection through the Bloch sphere origin. Summarizing these observations, we can state the following.

Theorem:

For all unital qubit channels, and all optimal signal ensembles $\{\,p_i\,,\,\rho_i\,\}$, the average density matrix $\sigma \;=\; \sum_i\, p_i\,\rho_i \;\equiv\; \frac{1}{2}\mathcal{I}_2$.

In Appendix A, it is shown that

$$\mathcal{D}\left(\,\rho\,\|\,\frac{1}{2}\mathcal{I}_2\,\right) \;=\; 1 \;-\; \mathcal{S}(\,\rho\,).$$

where $\mathcal{S}(\,\rho\,)$ is the von Neumann entropy of the density matrix ρ. Thus, our relation for the HSW channel capacity $\mathcal{C}_1$ becomes:

$$\mathcal{C}_1 \;=\; \sum_I\, p_i\,\mathcal{D}\left(\,\rho_i\,\|\,\frac{1}{2}\mathcal{I}_2\,\right) \;=\; 1 \;-\; \sum_i\, p_i\,\mathcal{S}(\,\rho_i\,).$$

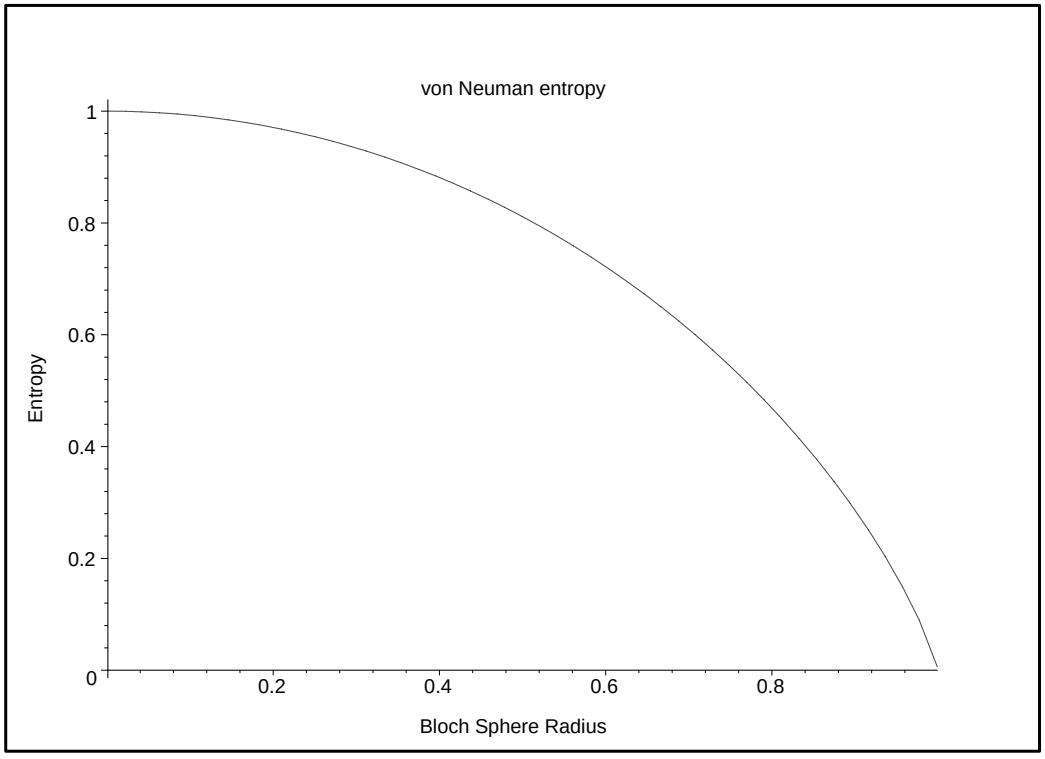

Figure 2.19: The von Neumann entropy $\mathcal{S}(\rho)$ for a single qubit ρ as a function of the Bloch sphere radius $r \in [0, 1]$.

To maximize $\mathcal{C}_1$, we seek to minimize the $\sum_i \mathcal{S}(\rho_i)$, subject to the constraint that the ρ_i satisfy $\sum_i p_i \rho_i = \frac{1}{2}\mathcal{I}_2$, for some set of a priori probabilities $\{p_i\}$. Recall that $\mathcal{S}(\rho) \equiv \mathcal{S}(r)$ is a strictly decreasing function of r, where r is the magnitude of the Bloch vector corresponding to ρ. (Please see Figure 19 for a plot of $\mathcal{S}(\rho) \equiv \mathcal{S}(r)$.)

Thus we seek to find a set of ρ_i which lie most distant, in terms of *Euclidean* distance in $\mathcal{R}^3$, from the ellipsoid origin, and for which a convex combination of these states equals the Bloch sphere origin.

Let us examine a few special cases. For the unital channel ellipsoid, consider the case where the major axis is unique in length, and has total length $2\,\lambda^{major\,axis}$. Let ρ_+ and ρ_- be the

states lying at the end of the major axis. By the symmetry of the ellipsoid, we have

$$\frac{1}{2}\,\rho_+ \;+\; \frac{1}{2}\,\rho_- \;=\; \frac{1}{2}\,\mathcal{I}_2.$$

Furthermore, the magnitude of the corresponding Bloch sphere vectors $r_+ \;=\; \|\,\vec{\mathcal{W}}_+\,\|$ and $r_- \;=\; \|\,\vec{\mathcal{W}}_-\,\|$ are equal, $r_+ \;=\; r_- \;=\; 1 \;-\; \left|\,\lambda^{major\;axis}\,\right|$.

Above, we use $|\cdots|$ around $\lambda^{major\;axis}$ because $\lambda^{major\;axis}$ can be a negative quantity in the King - Ruskai et al. formalism. Using this value of $r \;=\; r_+ \;=\; r_-$ yields for $\mathcal{C}_1$:

$$\mathcal{C}_1 \;=\; 1 \;-\; 2\,\left(\,\frac{1}{2}\,\mathcal{S}(r)\,\right) \;=\; 1 \;-\; \mathcal{S}\left(\,\left|\,\lambda^{major\;axis}\,\right|\,\right)\;.$$

If the major axis is not the unique axis of maximal length, then any set of convex probabilities and states $\{\,p_i\,,\;\rho_i\,\}$ such that the states lie on the major *surface* and $\sum_i\,p_i\,\rho_i \;\equiv\; \frac{1}{2}\,\mathcal{I}_2$ will suffice.

Thus we reach the same conclusion obtained by King and Ruskai in an earlier paper[25]. Summarizing, we can state the following.

Theorem:

The optimum output signalling states for unital qubit channels correspond to the minimum output von Neumann entropy states.

Furthermore, we can also conclude:

Theorem:

For unital qubit channels, the channel capacities consisting of signal state ensembles with two, three and four signalling states are equal. Furthermore, the optimum HSW channel capacity can be attained with a, possibly non-unique, pair of equiprobable ($p_1 \;=\; p_2 \;=\; \frac{1}{2}$) signalling states arranged opposite one another with respect to the Bloch sphere origin.

Proof:

Using the notation above, $C_2 = C_3 = C_4$. From the geometry of the centered channel ellipsoid, we can always use just two signalling states with the minimum output entropy to convexly reach $\frac{1}{2}\,\mathcal{I}_2$. Thus, utilizing more than two signaling states will not yield any channel capacity improvement beyond using two signalling states. The equiprobable nature of the two signalling states derives from the symmetry of the signalling states on the channel ellipsoid, in that one signalling state being the reflection of the other signalling state through the Bloch sphere origin means the states may be symmetrically added to yield an average state corresponding to the Bloch sphere origin. It is this reflection symmetry which makes the two signalling states equiprobable.

$\triangle$ - *End of Proof.*

The last three theorems were previously proven by King and Ruskai in section 2.3 of [25]. Here we have merely shown their results in the relative entropy picture.

VII.a The Depolarizing Channel

The depolarizing channel is a unital channel with $\{\, t_k = 0 \,\}$ and $\{\, \lambda_k = \frac{4\,x-1}{3} \,\}$, as discussed in more detail in Appendix C. The parameter $x \in [\,0\,,\,1\,]$. Using the analysis above, we can conclude that:

$$\mathcal{C}_1 = 1 - 2\left(\frac{1}{2}\,\mathcal{S}(r)\right) = 1 - \mathcal{S}\left(\left|\lambda^{major\,axis}\right|\right) = 1 - \mathcal{S}\left(\left|\frac{4\,x\,-\,1}{3}\right|\right)$$

$$= \frac{1 + \left|\frac{4\,x\,-\,1}{3}\right|}{2}\,\log_2\left(1 + \left|\frac{4\,x\,-\,1}{3}\right|\right) + \frac{1 - \left|\frac{4\,x\,-\,1}{3}\right|}{2}\,\log_2\left(1 - \left|\frac{4\,x\,-\,1}{3}\right|\right).$$

We plot $\mathcal{C}_1$ in Figure 20 below.

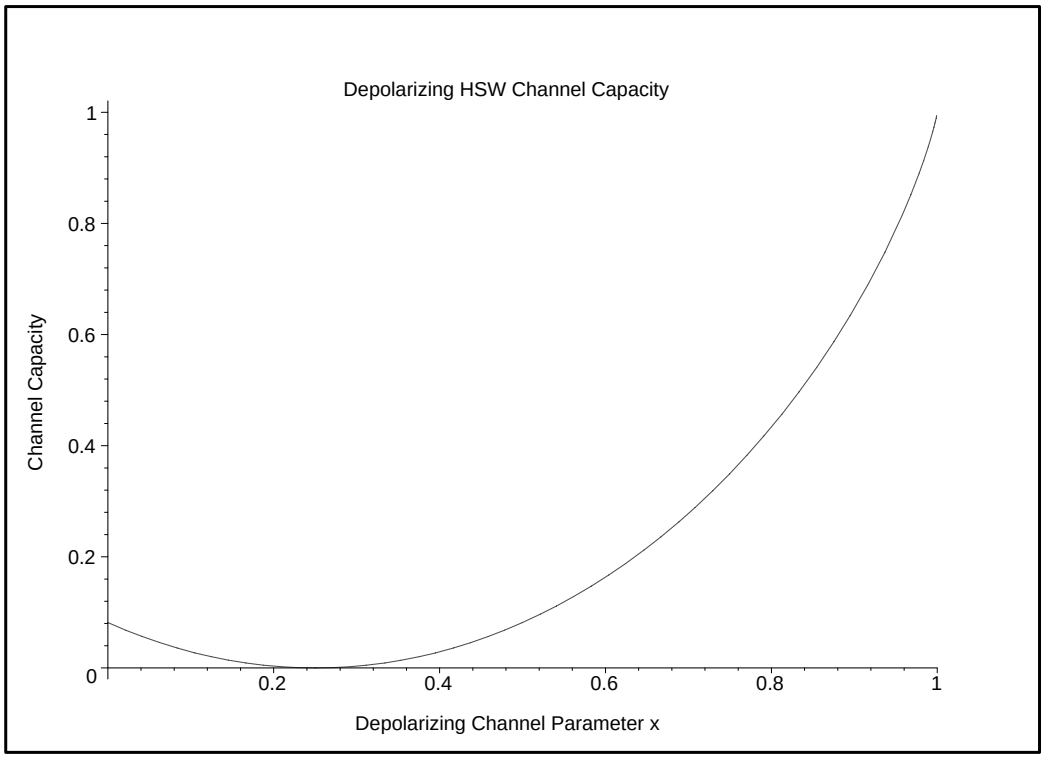

Figure 2.20: The Holevo-Schumacher-Westmoreland classical channel capacity for the Depolarizing channel as a function of the Depolarizing channel parameter x.

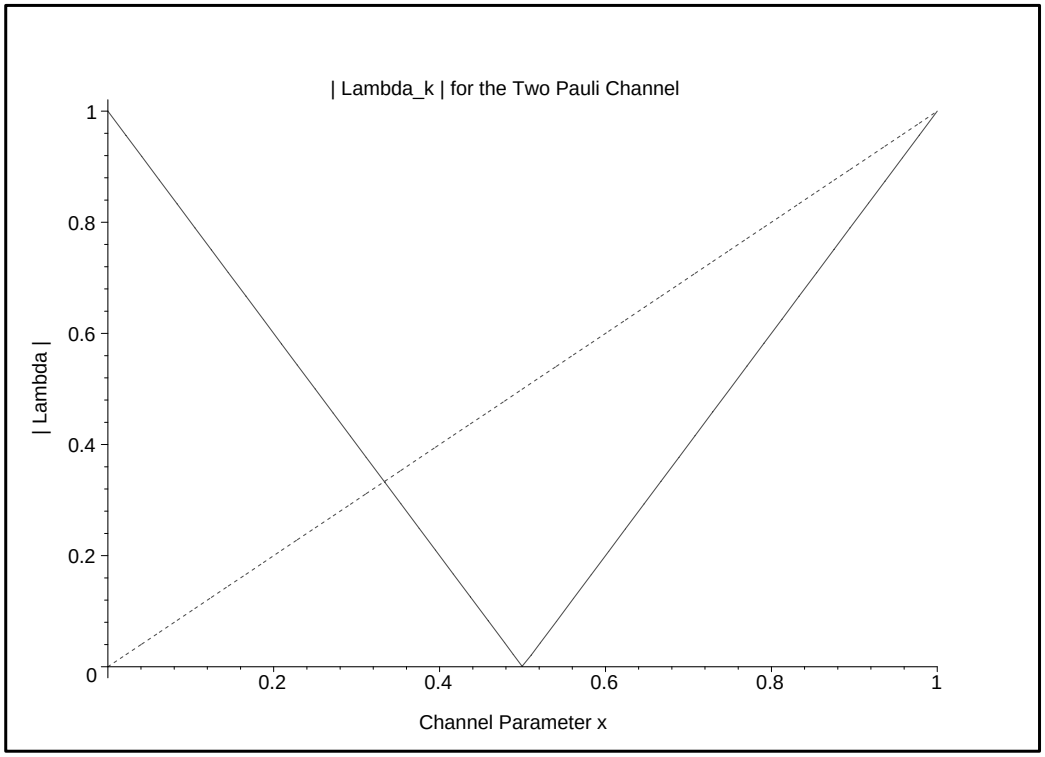

Figure 2.21: Calculating the length of the major axis of the channel ellipsoid for the Two Pauli channel as a function of the Two Pauli channel parameter x.

VII.b The Two Pauli Channel

The Two Pauli channel is a unital channel with $\{\, t_k \,=\, 0 \,\}$ and $\{\lambda_x \,=\, \lambda_y \,=\, x \,\}$, and $\{\lambda_z \,=\, 2\,x \,-\, 1 \,\}$, as discussed in more detail in Appendix C. The parameter $x \,\in\, [\,0\,,\,1\,]$. The determination of the major axis/surface is tricky due to the need to take into account the *absolute value* of the λ_k. We plot the absolute value of the λ_k in Figure 21 above. The dotted curve in Figure 21 corresponds to the absolute value of λ_x and λ_y. The V-shaped solid curve in Figure 21 corresponds to the absolute value of λ_z.

The intersection point occurs at $x \,=\, \frac{1}{3}$. Thus λ_z is the major axis for $x \,\leq\, \frac{1}{3}$ and the $\{\,\lambda_x\,,\,\lambda_y\,\}$ surface is the major axis surface for $x \,\geq\, \frac{1}{3}$. The Bloch sphere radius corresponding to the minimum entropy states is $1 \,-\, 2\,x$ for $x \,\leq\, \frac{1}{3}$ and x for $x \,\geq\, \frac{1}{3}$.

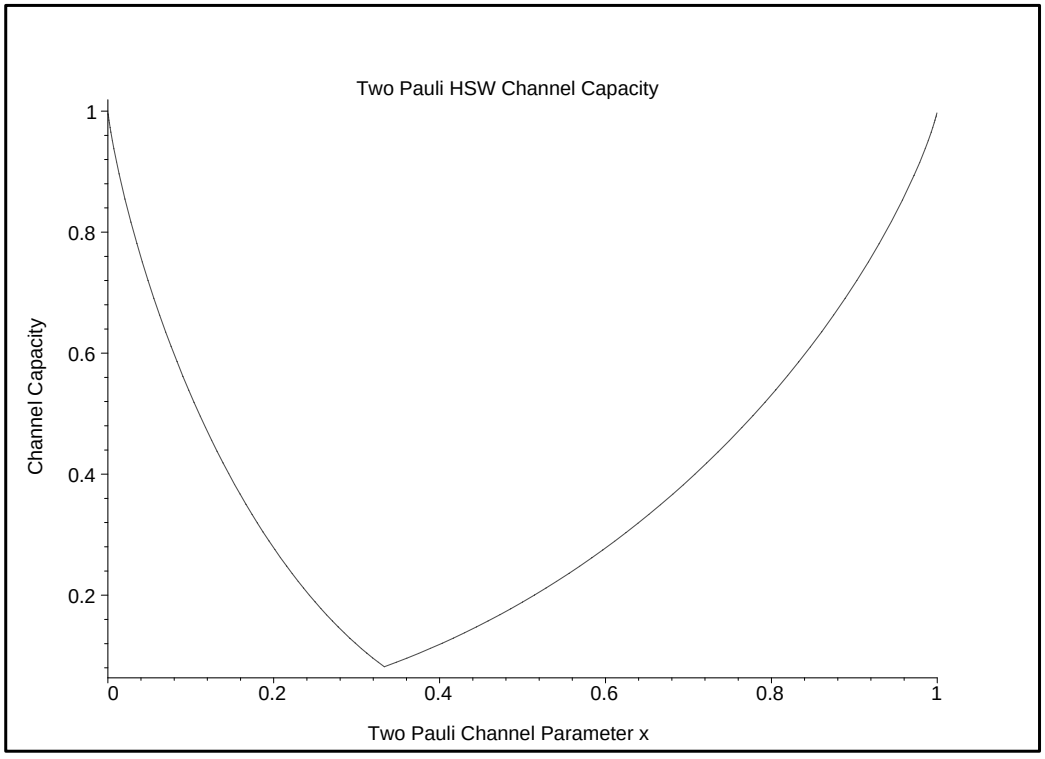

Figure 2.22: The Holevo-Schumacher-Westmoreland classical channel capacity for the Two Pauli channel as a function of the Two Pauli channel parameter x.

Using our analysis, we can conclude that for $x \leq \frac{1}{3}$, we have:

$$\mathcal{C}_1 = 1 - 2\left(\frac{1}{2}\mathcal{S}(r)\right) = 1 - \mathcal{S}(|\lambda_z|) = 1 - \mathcal{S}(1 - 2x)$$

$$= 1 + x\log_2(x) + (1-x)\log_2(1-x),$$

while for $x \geq \frac{1}{3}$, we have:

$$\mathcal{C}_1 = 1 - 2\left(\frac{1}{2}\mathcal{S}(r)\right) = 1 - \mathcal{S}(|\lambda_x|) = 1 - \mathcal{S}(x)$$

$$= \frac{1+x}{2}\log_2(1+x) + \frac{1-x}{2}\log_2(1-x).$$

We plot $\mathcal{C}_1$ in Figure 22 above, using the appropriate function in their allowed ranges of x.

Note the symmetry evident in the plots. Examining our graph of $\mathcal{C}_1$ in Figure 22, one can see that for $0 \leq \alpha \leq \frac{1}{3}$, we have $\mathcal{C}_1(\frac{1}{3} - \alpha) \equiv \mathcal{C}_1(\frac{1}{3} + 2\alpha)$. This symmetry is also readily seen from the relations for $\mathcal{C}_1$ in the two allowed ranges of x (less than and greater than $\frac{1}{3}$).

For $x \leq \frac{1}{3}$, setting $x = \frac{1}{3} - \alpha$,

$$\mathcal{C}_1^-(\alpha) = 1 + \frac{1-3\alpha}{3} \log_2\left(\frac{1-3\alpha}{3}\right) + \frac{2+3\alpha}{3}\log_2\left(\frac{2+3\alpha}{3}\right).$$

For $x \geq \frac{1}{3}$, setting $x = \frac{1}{3} + 2\alpha$,

$$\mathcal{C}_1^+(\alpha) = \frac{4+6\alpha}{6} \log_2\left(\frac{4+6\alpha}{3}\right) + \frac{2-6\alpha}{6}\log_2\left(\frac{2-6\alpha}{3}\right)$$

$$= \left(\frac{4+6\alpha}{6} + \frac{2-6\alpha}{6}\right) + \frac{4+6\alpha}{6} \log_2\left(\frac{2+3\alpha}{3}\right) + \frac{2-6\alpha}{6}\log_2\left(\frac{1-3\alpha}{3}\right)$$

$$= 1 + \frac{2+3\alpha}{3} \log_2\left(\frac{2+3\alpha}{3}\right) + \frac{1-3\alpha}{3}\log_2\left(\frac{1-3\alpha}{3}\right) = \mathcal{C}_1^-(\alpha).$$

VIII Non-Unital Channels

Non-unital channels are generically more difficult to analyze due to the fact that one or more of the $\{ t_k \}$ can be non-zero. This allows the average density matrix $\rho = \sum_i p_i \rho_i$ for an optimal signal ensemble $\{ p_i, \rho_i \}$ to move away from the Bloch sphere origin $\rho = \frac{1}{2}\mathcal{I}_2 \equiv \vec{\mathcal{V}} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$. However, there still remains the symmetry present in the qubit form of the relative entropy formula, namely that $\mathcal{D}(\rho \| \phi) = \mathcal{D}(\vec{\mathcal{W}} \| \vec{\mathcal{V}}) = f(r, q, \theta)$, where $r = \| \vec{\mathcal{W}} \|$, $q = \| \vec{\mathcal{V}} \|$, and θ is the angle between $\vec{\mathcal{W}}$ and $\vec{\mathcal{V}}$. The fact that the qubit relative entropy depends only on r, q, and θ yields a symmetry which can be used to advantage in analyzing non-unital channels, as our last example will demonstrate.

VIII.a The Amplitude Damping Channel

The amplitude damping channel is a non-unital channel with $\{\, t_x = t_y = 0 \,\}$ and $\{\, t_z = 1 - \xi \,\}$. The λ_k are $\{\lambda_x = \lambda_y = \sqrt{\xi}\,\}$, and $\{\lambda_z = \xi\,\}$, where ξ is the channel parameter, $\xi \in [\,0\,,\,1\,]$. The amplitude damping channel is discussed in more detail in Appendix C. The determination of the major axis/surface reduces to an analysis in either the X-Y or X-Z Bloch sphere *plane* because of symmetries of the channel ellipsoid *and* the relative entropy formula for qubit density matrices. Since the relative entropy formula depends only on the r, q and θ quantities which were defined above, by examining contour *curves* of relative entropy in the X-Z plane, we can create a *surface* of constant relative entropy in the three dimensional X-Y-Z Bloch sphere space by the solid of revolution technique. That is, we shall revolve our X-Z contour curves about the axis of symmetry, here the Z-axis. Now the channel ellipsoid in this case is also rotationally symmetric about the Z-axis, because $t_x = t_y = 0$ and $\lambda_x = \lambda_y$. Thus optimum signal *points* (points on the channel ellipsoid surface which have maximal relative entropy distance from the average signal density matrix), in the X-Z plane, will become *circles* of optimal signals in the full three dimensional Bloch sphere picture after the revolution about the Z - axis is completed. Therefore, due to the simultaneous rotational symmetry about the Bloch sphere Z axis of the relative entropy formula (for qubits) and the channel ellipsoid, a full three dimensional analysis of the amplitude damping channel reduces to a much easier, yet equivalent, two dimensional analysis in the Bloch X-Z plane.

To illustrate these ideas, we take a specific instance of the amplitude damping channel with $\xi = 0.36$. Then $\{\, t_x = t_y = 0 \,\}$ and $\{\, t_z = 0.64 \,\}$. The λ_k are $\{\lambda_x = \lambda_y = 0.6\,\}$, and $\{\lambda_z = 0.36\,\}$. In this case $\mathcal{C}_1 = 0.3600$ is achieved with two equiprobable signalling states. The optimum average density matrix has Bloch vector $\vec{\mathcal{V}} = \begin{bmatrix} 0 \\ \\ 0.7126 \end{bmatrix}$, and is shown as an asterisk (*) in Figures 23 and 24 below.

In the first plot, Figure 23, we show the X(horizontal)-Z(vertical) Bloch sphere plane. The outer bold dotted ring is the pure state boundary, with Bloch vector magnitude equal to

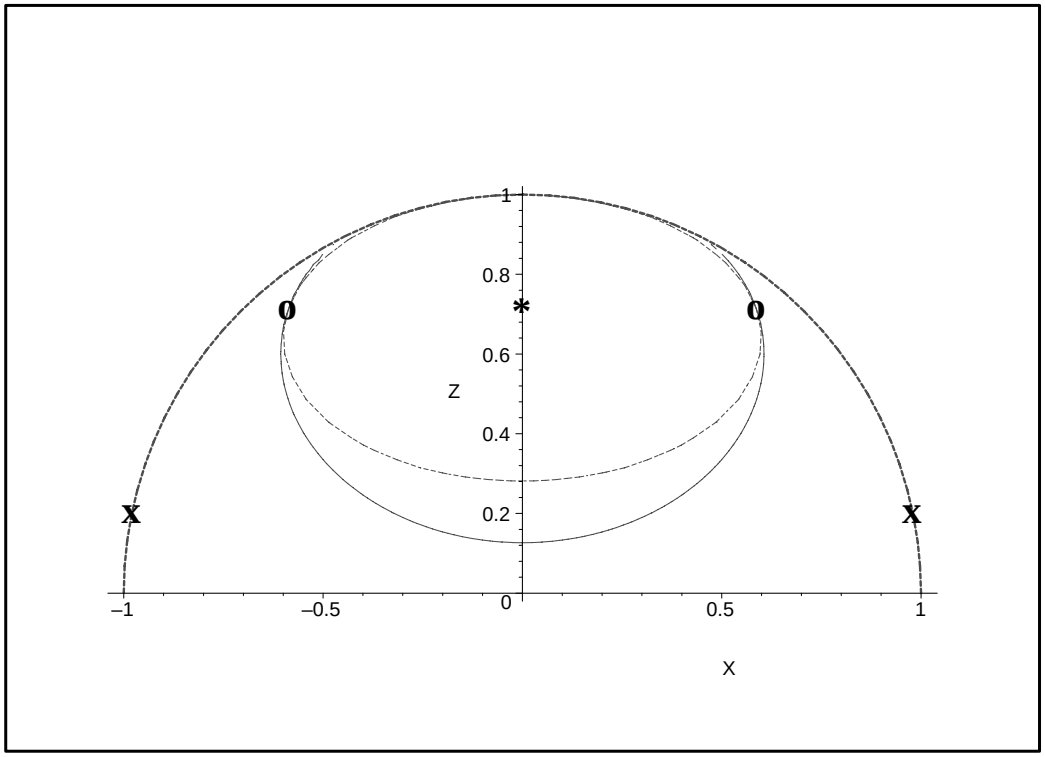

Figure 2.23: The intersection in the Bloch sphere X-Z plane of the Amplitude Damping channel ellipsoid (the inner dashed curve) and the optimum Relative Entropy contour (the solid curve). The two optimum input signal states (on the outer bold dashed Bloch sphere boundary curve) are shown as **X**. The two optimum output signal states (on the channel ellipsoid *and* the optimum Relative Entropy contour curve) are shown as **O**.

one. The inner dashed circle is the channel ellipsoid. The middle solid contour is the curve of constant relative entropy, equal to 0.3600, and centered at $\vec{\mathcal{V}}$ (*). The relative entropy contour in the X-Z plane contacts the channel ellipsoid at two *symmetrical* points, indicated in the plot as **O**. Note that the two contact points of the relative entropy contour and the channel ellipsoid contour (the two **O** points), and the location of $\vec{\mathcal{V}}$ (*), all lie on a perfectly horizontal line. The fact that the line is horizontal is due to the fact that the two optimum signalling states in the X-Z plane are symmetric about the Z axis. The point $\vec{\mathcal{V}}$ (*) is simply the two optimal output signal points average. The corresponding optimal input signals are shown as **X**'s on the outer bold dotted pure state boundary semicircular curve.

Note that the optimum input signalling states are nonorthogonal. Furthermore, this analysis

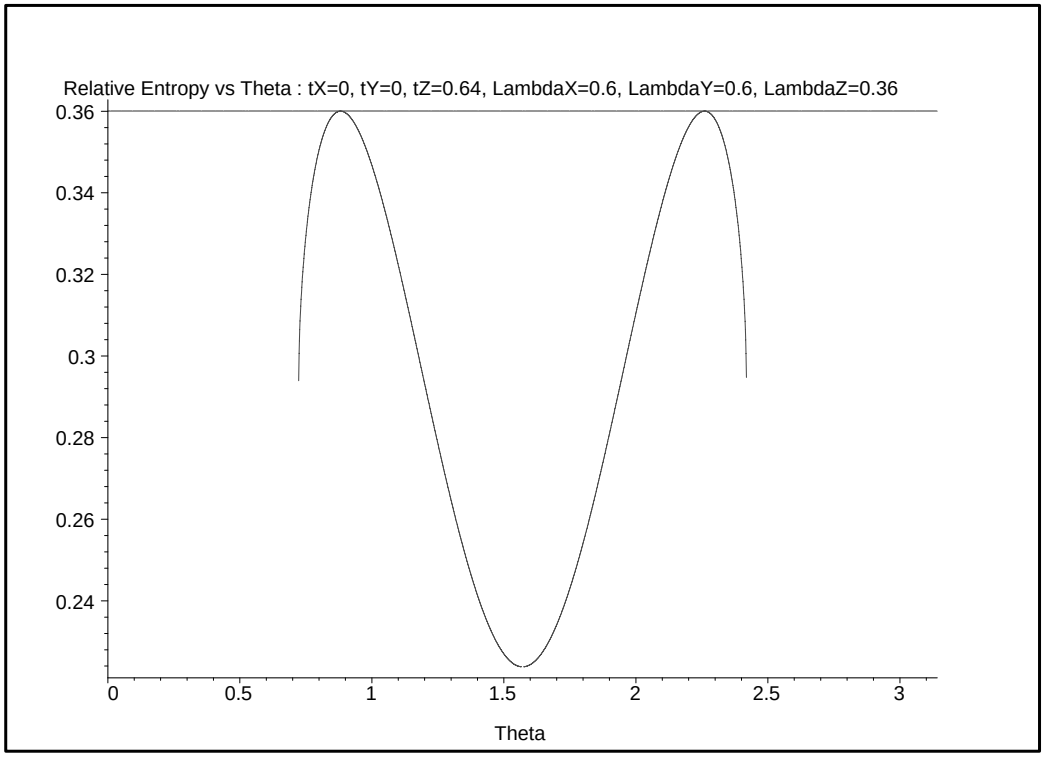

Figure 2.24: The change in $\mathcal{D}(\ \rho \parallel \phi \equiv * \ )$ as we move ρ around the channel ellipsoid. The angle Theta is with respect to the Bloch sphere origin.

tells us that $C_2 = C_3 = C_4$. For the amplitude damping channel, there is no advantage to using more than two signals in the optimum signalling ensemble.

Figure 24 is a picture similar to those we have done for the planar channels we previously examined. In Figure 24 we plot the magnitude of the relative entropy as one moves around the channel ellipsoid in the X-Z plane. The angle θ is with respect to the Bloch sphere origin (ie: the X-Z plane origin).

Thus the rotational symmetry about the Z-axis of the relative entropy formula, coupled with the same Z - axis rotational symmetry of the amplitude damping channel ellipsoid, yields a complete understanding of the behavior of the amplitude damping channel with just a simple two dimensional analysis.

IX Summary and Conclusions

In this chapter, we have derived a formula for the relative entropy of two single qubit density matrices. By combining our relative entropy formula with the King-Ruskai et al. ellipsoid picture of qubit channels, we can use the Schumacher-Westmoreland relative entropy approach to classical HSW channel capacity to analyze unital and non-unital single qubit channels in detail.

The following observation also emerges from the examples and analyses above. In numerical simulations by this author and others, it was noted that the a priori probabilities of the optimum signalling states for non-unital qubit channels were in general, approximately, but not exactly, equal. For example, consider the case of linear channels, where the optimum HSW channel capacity is achieved with two signalling states. In our first linear channel example, one signalling state had an a priori probability of 0.5156 and the other signalling state had an a priori probability of 0.4844. Similarly, in our second linear channel example, the respective a priori probabilities were 0.5267 and 0.4733. These asymmetries in the a priori probabilities are due to the fact that $\mathcal{D}$ is not purely a radial function of distance from $\vec{\mathcal{V}}_{optimum}$. The relative entropy contours shown in Figures 2.2 through 2.11 are moderately, but not exactly, circular about $\vec{\mathcal{V}}_{optimum}$. This slight radial asymmetry leads to a priori signal probabilities that are approximately, but not exactly, equal. Thus, a graphical estimate of the a priori signal probabilities can be made by observing the degree of asymmetry of the optimum relative entropy contour about $\vec{\mathcal{V}}_{optimum}$.

In conclusion, the analysis above yields a geometric picture which we hope will lead to future insights into the transmission of classical information over single qubit channels.

X Appendix A: The Bloch Sphere Relative Entropy Formula

The relative entropy of two density matrices ϱ and ψ is defined to be

$$\mathcal{D}(\ \varrho\,||\,\psi\ )\ =\ Tr[\ \varrho\,(\ \log_2(\varrho)\ -\ \log_2(\psi)\ )\ ].$$

Our main interest is when both ϱ and ψ are qubit density operators. In that case, ϱ and ψ can be written using the Bloch sphere representation.

$$\varrho\ =\ \frac{1}{2}\ \left(\mathcal{I}_2\ +\ \vec{\mathcal{W}}\bullet\vec{\sigma}\right)\qquad\qquad \psi\ =\ \frac{1}{2}\ \left(\mathcal{I}_2\ +\ \vec{\mathcal{V}}\bullet\vec{\sigma}\right).$$

To simplify notation below, we define

$$r\ =\ \sqrt{\vec{\mathcal{W}}\bullet\vec{\mathcal{W}}}\qquad and\qquad q\ =\ \sqrt{\vec{\mathcal{V}}\bullet\vec{\mathcal{V}}}.$$

We shall also define $\cos(\theta)$ as:

$$\cos(\theta)\ \ =\ \ \frac{\vec{\mathcal{W}}\bullet\vec{\mathcal{V}}}{r\ q},$$

where r and q are as above.

The symbol $\vec{\sigma}$ means the vector of 2 x 2 Pauli matrices

$$\vec{\sigma}\ =\ \begin{bmatrix}\sigma_x\\ \sigma_y\\ \sigma_z\end{bmatrix}\qquad where\qquad \sigma_x\ =\ \begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix},\qquad \sigma_y\ =\ \begin{bmatrix}0 & -i\\ i & 0\end{bmatrix},\qquad \sigma_z\ =\ \begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}.$$

The Bloch vectors $\vec{\mathcal{W}}$ and $\vec{\mathcal{V}}$ are real, three dimensional vectors which have magnitude equal to one when representing a pure state density matrix, and magnitude less than one for a mixed (non-pure) density matrix.

The density matrices for ϱ and ψ in terms of their Bloch vectors are:

$$\varrho \;=\; \begin{bmatrix} \frac{1}{2} + \frac{1}{2}\,w_3 & \frac{1}{2}\,w_1 - \frac{1}{2}\,i w_2 \\ \frac{1}{2}\,w_1 + \frac{1}{2}\,i w_2 & \frac{1}{2} - \frac{1}{2}\,w_3 \end{bmatrix}.$$

$$\psi \;=\; \begin{bmatrix} \frac{1}{2} + \frac{1}{2}\,v_3 & \frac{1}{2}\,v_1 - \frac{1}{2}\,i v_2 \\ \frac{1}{2}\,v_1 + \frac{1}{2}\,i v_2 & \frac{1}{2} - \frac{1}{2}\,v_3 \end{bmatrix}.$$

We shall prove the following formula in two ways, an algebraic proof and a brute force proof. We conclude Appendix A with some alternate representations of this formula.

$$\mathcal{D}(\,\varrho\,\|\,\psi\,) \;=\; \mathcal{D}_1 \;-\; \mathcal{D}_2 \;=$$

$$\frac{1}{2}\log_2\left(1 \;-\; r^2\right) \;+\; \frac{r}{2}\log_2\left(\frac{1 \;+\; r}{1 \;-\; r}\right) \;-\; \frac{1}{2}\,\log_2\left(1 \;-\; q^2\,\right) \;-\; \frac{\vec{\mathcal{W}} \bullet \vec{\mathcal{V}}}{2\,q}\,\log_2\left(\frac{1 \;+\; q}{1 \;-\; q}\right)$$

$$= \frac{1}{2}\log_2\left(1 \;-\; r^2\right) \;+\; \frac{r}{2}\log_2\left(\frac{1 \;+\; r}{1 \;-\; r}\right) \;-\; \frac{1}{2}\,\log_2\left(1 \;-\; q^2\,\right) \;-\; \frac{r\,\cos(\theta)}{2}\,\log_2\left(\frac{1 \;+\; q}{1 \;-\; q}\right)$$

where θ is the angle between $\vec{\mathcal{W}}$ and $\vec{\mathcal{V}}$.

X.a Proof I: The Algebraic Proof

$$\mathcal{D}(\,\varrho\,||\,\psi\,) \;=\; Tr[\,\varrho\,(\,\log_2(\varrho) \;-\; \log_2(\psi)\,)\,].$$

Recall the following Taylor series, valid for $\|\,x\,\| \;\leq\; 1$.

$$\ln(\,1 \;+\; x\,) \;=\; -\sum_{n=1}^{\infty}\,\frac{(\,-x\,)^n}{n} \;=\; x \;-\; \frac{x^2}{2} \;+\; \frac{x^3}{3} \;-\; \frac{x^4}{4} \;+\; \frac{x^5}{5} \;-\; \frac{x^6}{6} \;+\; \frac{x^7}{7} \;-\; \cdots$$

$$\ln(\,1 \;-\; x\,) \;=\; -\sum_{n=1}^{\infty}\,\frac{x^n}{n} \;=\; -\,x \;-\; \frac{x^2}{2} \;-\; \frac{x^3}{3} \;-\; \frac{x^4}{4} \;-\; \frac{x^5}{5} \;-\; \frac{x^6}{6} \;-\; \frac{x^7}{7} \;-\; \cdots$$

Combining these two Taylor series yields another Taylor expansion we shall be interested in:

$$\frac{1}{2}\,\{\,\ln(\,1\,+\,x\,)\,-\,\ln(\,1\,-\,x\,)\,\}\,=\,\frac{1}{2}\,\ln\left(\,\frac{1\,+\,x}{1\,-\,x}\,\right)$$

$$=\,-\,\sum_{n=1}^{\infty}\,\frac{(-x)^n}{n}\quad-\quad\left(\,-\,\sum_{n=1}^{\infty}\,\frac{x^n}{n}\,\right)\,=\,x\,+\,\frac{x^3}{3}\,+\,\frac{x^5}{5}\,+\,\frac{x^7}{7}\,+\,\frac{x^9}{9}\,+\,\cdots$$

A different combination of the first two Taylor series above yields yet another Taylor expansion we shall be interested in:

$$\frac{1}{2}\,\{\,\ln(\,1\,+\,x\,)\,+\,\ln(\,1\,-\,x\,)\,\}\,=\,\frac{1}{2}\,\ln\Big[\,1\,-\,x^2\,\Big]$$

$$=\,-\,\sum_{n=1}^{\infty}\,\frac{(-x)^n}{n}\quad+\quad\left(\,-\,\sum_{n=1}^{\infty}\,\frac{x^n}{n}\,\right)\,=\,-\,\frac{x^2}{2}\,-\,\frac{x^4}{4}\,-\,\frac{x^6}{6}\,-\,\frac{x^8}{8}\,-\,\cdots$$

Consider $\log(\,\varrho\,)$ with the Bloch sphere representation for ϱ.

$$\varrho\,=\,\frac{1}{2}\,\Big(\mathcal{I}_2\,+\,\vec{\mathcal{W}}\bullet\vec{\sigma}\Big)\,.$$

We obtain, using the expansion given above for $\log(\,1\,+\,x\,)$,

$$\log(\,\varrho\,)\,=\,\log\Big[\,\frac{1}{2}\,\Big(\,\mathcal{I}_2\,+\,\vec{\mathcal{W}}\bullet\vec{\sigma}\,\Big)\,\Big]\,=\,\log\Big[\,\frac{1}{2}\,\Big]\,+\,\log\Big[\,\mathcal{I}_2\,+\,\vec{\mathcal{W}}\bullet\vec{\sigma}\,\Big]$$

$$=\,\log\Big[\,\frac{1}{2}\,\Big]\,-\,\sum_{n=1}^{\infty}\,\frac{\Big(\,-\,\vec{\mathcal{W}}\bullet\vec{\sigma}\,\Big)^n}{n}.$$

Recall that $\Big(\,\vec{\mathcal{W}}\bullet\vec{\sigma}\,\Big)^2\;=\;r^2$, where $r\;=\;\sqrt{\vec{\mathcal{W}}\bullet\vec{\mathcal{W}}}$. Thus we have for even n, $\Big(\,\vec{\mathcal{W}}\bullet\vec{\sigma}\,\Big)^n\;=\;r^n$, while for odd n we have $\Big(\,\vec{\mathcal{W}}\bullet\vec{\sigma}\,\Big)^n\;=\;r^{n-1}\;\;\vec{\mathcal{W}}\bullet\vec{\sigma}$. The expression for $\log(\,\varrho\,)$ then becomes

$$\log(\,\varrho\,)\,=\,\log\Big[\,\frac{1}{2}\,\Big]\quad-\quad\sum_{n=1}^{\infty}\,\frac{\Big(\,-\,\vec{\mathcal{W}}\bullet\vec{\sigma}\,\Big)^n}{n}$$

$$=\,\log\Big[\,\frac{1}{2}\,\Big]\quad+\quad\vec{\mathcal{W}}\bullet\vec{\sigma}\,-\,\frac{r^2}{2}\,+\,\frac{r^2}{3}\,\vec{\mathcal{W}}\bullet\vec{\sigma}\,-\,\frac{r^4}{4}\,+\,\frac{r^4}{5}\,\vec{\mathcal{W}}\bullet\vec{\sigma}\,-\,\frac{r^6}{6}\,+\,\frac{r^6}{7}\,\vec{\mathcal{W}}\bullet\vec{\sigma}\,-\,\cdots$$

$$= \log\left[\frac{1}{2}\right] + \frac{\vec{\mathcal{W}}\bullet\vec{\sigma}}{r}\left(r + \frac{r^3}{3} + \frac{r^5}{5} + \frac{r^7}{7} + \cdots\right)$$

$$+ \left(-\frac{r^2}{2} - \frac{r^4}{4} - \frac{r^6}{6} - \frac{r^8}{8} - \frac{r^{10}}{10} - \cdots\right)$$

$$= \log\left[\frac{1}{2}\right] + \frac{\vec{\mathcal{W}}\bullet\vec{\sigma}}{2\,r}\log\left[\frac{1+r}{1-r}\right] + \frac{1}{2}\log\left[1 - r^2\right].$$

To evaluate $Tr[\,\varrho\,\log(\,\varrho\,)\,]$ we again use the Bloch sphere representation for ϱ.

$$\varrho = \frac{1}{2}\left(\mathcal{I}_2 + \vec{\mathcal{W}}\bullet\vec{\sigma}\right).$$

We write

$$Tr[\,\varrho\,\log(\,\varrho\,)\,] = \frac{1}{2}Tr[\,\mathcal{I}_2 \bullet \log(\,\varrho\,)\,] + \frac{1}{2}Tr\left[\left(\vec{\mathcal{W}}\bullet\vec{\sigma}\right)\log(\,\varrho\,)\right].$$

Using our results above,

$$\frac{1}{2}Tr[\,\mathcal{I}_2 \bullet \log(\,\varrho\,)\,] = \log\left[\frac{1}{2}\right] + \frac{1}{2}\log\left[1 - r^2\right].$$

since $Tr[\,\mathcal{I}_2\,] = 2$ and $Tr[\,\sigma_x\,] = Tr[\,\sigma_y\,] = Tr[\,\sigma_z\,] = 0$. Similarly,

$$Tr\left[\left(\vec{\mathcal{W}}\bullet\vec{\sigma}\right)\log(\,\varrho\,)\right] = \frac{\left(\vec{\mathcal{W}}\bullet\vec{\sigma}\right)^2}{r}\log\left[\frac{1+r}{1-r}\right] = r\,\log\left[\frac{1+r}{1-r}\right].$$

where we again used the fact $Tr[\,\mathcal{I}_2\,] = 2$ and $Tr[\,\sigma_x\,] = Tr[\,\sigma_y\,] = Tr[\,\sigma_z\,] = 0$. Putting all the pieces together yields:

$$Tr[\,\varrho\,\log(\,\varrho\,)\,] = \frac{1}{2}Tr[\,\mathcal{I}_2 \bullet \log(\,\varrho\,)\,] + \frac{1}{2}Tr\left[\left(\vec{\mathcal{W}}\bullet\vec{\sigma}\right)\log(\,\varrho\,)\right]$$

$$= \log\left[\frac{1}{2}\right] + \frac{1}{2}\log\left[1 - r^2\right] + \frac{r}{2}\log\left[\frac{1+r}{1-r}\right].$$

To evaluate $Tr[\,\varrho\,\log(\,\psi\,)\,]$, we follow a similar path and use the Bloch sphere representation for ψ of

$$\psi = \frac{1}{2}\left(\mathcal{I}_2 + \vec{\mathcal{V}}\bullet\vec{\sigma}\right).$$

The expression for $\log(\psi)$ then becomes

$$\log(\psi) = \log\left[\frac{1}{2}\right] + \frac{\vec{\mathcal{V}} \bullet \vec{\sigma}}{2\,q} \log\left[\frac{1+q}{1-q}\right] + \frac{1}{2}\log\left[1-q^2\right].$$

Using our results above,

$$\frac{1}{2}\,Tr[\,\mathcal{I}_2 \bullet \log(\psi)\,] = \log\left[\frac{1}{2}\right] + \frac{1}{2}\log\left[1-q^2\right] = -\log[2] + \frac{1}{2}\log\left[1-q^2\right].$$

$$Tr\left[\left(\vec{\mathcal{W}} \bullet \vec{\sigma}\right)\log(\psi)\right] = \frac{\vec{\mathcal{W}} \bullet \vec{\mathcal{V}}}{q}\log\left[\frac{1+q}{1-q}\right] = r\,\cos(\theta)\log\left[\frac{1+q}{1-q}\right],$$

where we again used the fact $Tr[\,\mathcal{I}_2\,] = 2$ and $Tr[\,\sigma_x\,] = Tr[\,\sigma_y\,] = Tr[\,\sigma_z\,] = 0$. We also used the fact that

$$\left(\vec{\mathcal{V}} \bullet \vec{\sigma}\right)\left(\vec{\mathcal{W}} \bullet \vec{\sigma}\right) = \left(\vec{\mathcal{V}} \bullet \vec{\mathcal{W}}\right)\mathcal{I}_2 + \left(\vec{\mathcal{V}} \times \vec{\mathcal{W}}\right) \bullet \vec{\sigma}$$

and therefore

$$Tr\left[\left(\vec{\mathcal{V}} \bullet \vec{\sigma}\right)\left(\vec{\mathcal{W}} \bullet \vec{\sigma}\right)\right] = Tr\left[\left(\vec{\mathcal{V}} \bullet \vec{\mathcal{W}}\right)\mathcal{I}_2\right] + Tr\left[\left(\vec{\mathcal{V}} \times \vec{\mathcal{W}}\right) \bullet \vec{\sigma}\right]$$

$$= \left(\vec{\mathcal{V}} \bullet \vec{\mathcal{W}}\right) Tr[\,\mathcal{I}_2\,] + \left(\vec{\mathcal{V}} \times \vec{\mathcal{W}}\right) \bullet Tr[\,\vec{\sigma}\,] = 2\,\vec{\mathcal{V}} \bullet \vec{\mathcal{W}}.$$

Assembling the pieces:

$$Tr[\,\varrho\,\log(\psi)\,] = \frac{1}{2}\,Tr[\,\mathcal{I}_2 \bullet \log(\psi)\,] + \frac{1}{2}\,Tr\left[\left(\vec{\mathcal{W}} \bullet \vec{\sigma}\right)\log(\psi)\right]$$

$$= \log\left[\frac{1}{2}\right] + \frac{1}{2}\log\left[1-q^2\right] + \frac{r}{2}\cos(\theta)\log\left[\frac{1+q}{1-q}\right].$$

Using these pieces, we obtain our final formula:

$$\mathcal{D}(\,\varrho\,||\,\psi\,) = Tr[\,\varrho\,(\,\log_2(\varrho) - \log_2(\psi)\,)\,]$$

$$= \log_2\left[\frac{1}{2}\right] + \frac{1}{2}\log_2\left[1-r^2\right] + \frac{r}{2}\log_2\left[\frac{1+r}{1-r}\right]$$

$$- \log_2 \left[\frac{1}{2} \right] - \frac{1}{2} \log_2 \left[1 - q^2 \right] - \frac{r}{2} \cos(\theta) \log_2 \left[\frac{1+q}{1-q} \right]$$

$$= \frac{1}{2} \log_2 \left[1 - r^2 \right] + \frac{r}{2} \log_2 \left[\frac{1+r}{1-r} \right] - \frac{1}{2} \log_2 \left[1 - q^2 \right] - \frac{r}{2} \cos(\theta) \log_2 \left[\frac{1+q}{1-q} \right],$$

which is our desired formula.

$\triangle$ - *End of Proof I.*

X.b Proof II: The Brute Force Proof

The density matrices for ϱ and ψ in terms of their Bloch vectors are:

$$\varrho = \begin{bmatrix} \frac{1}{2} + \frac{1}{2} w_3 & \frac{1}{2} w_1 - \frac{1}{2} i w_2 \\ \frac{1}{2} w_1 + \frac{1}{2} i w_2 & \frac{1}{2} - \frac{1}{2} w_3 \end{bmatrix}$$

$$\psi = \begin{bmatrix} \frac{1}{2} + \frac{1}{2} v_3 & \frac{1}{2} v_1 - \frac{1}{2} i v_2 \\ \frac{1}{2} v_1 + \frac{1}{2} i v_2 & \frac{1}{2} - \frac{1}{2} v_3 \end{bmatrix}.$$

The eigenvalues of these two density matrices are:

$$\lambda_\varrho^{(1)} = \frac{1}{2} + \frac{1}{2} \sqrt{{w_2}^2 + {w_3}^2 + {w_1}^2} = \frac{1+r}{2}.$$

$$\lambda_\varrho^{(2)} = \frac{1}{2} - \frac{1}{2} \sqrt{{w_2}^2 + {w_3}^2 + {w_1}^2} = \frac{1-r}{2}.$$

$$\lambda_\psi^{(1)} = \frac{1}{2} + \frac{1}{2} \sqrt{{v_2}^2 + {v_3}^2 + {v_1}^2} = \frac{1+q}{2}.$$

$$\lambda_\psi^{(2)} = \frac{1}{2} - \frac{1}{2} \sqrt{{v_2}^2 + {v_3}^2 + {v_1}^2} = \frac{1-q}{2}.$$

We shall also be interested in the two eigenvectors of ψ. These are:

$$|e_1\rangle \;=\; N_1 \left[\begin{array}{c} 1 \\ -\frac{-2\left(\frac{1}{2}+\frac{1}{2}\sqrt{{w_2}^2+{w_3}^2+{w_1}^2}\right)w_1 - 2\,i\left(\frac{1}{2}+\frac{1}{2}\sqrt{{w_2}^2+{w_3}^2+{w_1}^2}\right)w_2 + w_1 + i w_2 + w_3\; w_1 + i w_3\; w_2}{{w_1}^2+{w_2}^2} \end{array} \right],$$

where N_1 is the normalization constant given below.

$$N_1 \;=\; \sqrt{2\,\frac{{w_1}^2 + \sqrt{{w_2}^2+{w_3}^2+{w_1}^2}w_3 + {w_2}^2 + {w_3}^2}{{w_1}^2+{w_2}^2}}.$$

Similarly,

$$|e_2\rangle \;=\; N_2 \left[\begin{array}{c} \frac{2\left(\frac{1}{2}-\frac{1}{2}\sqrt{{w_2}^2+{w_3}^2+{w_1}^2}\right)w_1 - 2\,i\left(\frac{1}{2}-\frac{1}{2}\sqrt{{w_2}^2+{w_3}^2+{w_1}^2}\right)w_2 - w_1 + i w_2 + w_3\; w_1 - i w_3\; w_2}{{w_1}^2+{w_2}^2} \\ 1 \end{array} \right].$$

$$N_2 \;=\; \sqrt{2\,\frac{{w_1}^2 + \sqrt{{w_2}^2+{w_3}^2+{w_1}^2}w_3 + {w_2}^2 + {w_3}^2}{{w_1}^2+{w_2}^2}}.$$

We wish to derive a formula for $\mathcal{D}(\varrho \,\|\, \psi)$ in terms of the Bloch sphere vectors $\vec{\mathcal{W}}$ and $\vec{\mathcal{V}}$. We do this by breaking $\mathcal{D}(\varrho \,\|\, \psi)$ up into two terms, $\mathcal{D}_1$ and $\mathcal{D}_2$.

$$\mathcal{D}(\varrho \,\|\, \psi) \;=\; \mathcal{D}_1 \;-\; \mathcal{D}_2.$$

We expand $\mathcal{D}_1$ using our knowledge of the eigenvalues of ϱ.

$$\begin{aligned} \mathcal{D}_1 \;&=\; Tr[\,\varrho\;\log_2(\varrho)\,] \\ &=\; \lambda_\varrho^{(1)}\;\log_2(\lambda_\varrho^{(1)}) \;+\; \lambda_\varrho^{(2)}\;\log_2(\lambda_\varrho^{(2)}) \\ &=\; \left(\frac{1\;+\;r}{2}\right)\log_2\left(\frac{1\;+\;r}{2}\right) \;+\; \left(\frac{1\;-\;r}{2}\right)\log_2\left(\frac{1\;-\;r}{2}\right) \\ &=\; -\;1\;+\;\left(\frac{1\;+\;r}{2}\right)\log_2\left(1\;+\;r\right) \;+\; \left(\frac{1\;-\;r}{2}\right)\log_2\left(1\;-\;r\right). \end{aligned}$$

One notes that $\mathcal{D}_1 \;=\; -\mathcal{S}(\varrho)$, where $\mathcal{S}(\varrho)$ is the von Neumann entropy of the density matrix ϱ. The second term, $\mathcal{D}_2$, is $\mathcal{D}_2 \;=\; Tr[\,\varrho\;\log_2(\psi)\,]$. We evaluate $\mathcal{D}_2$ in the basis which

diagonalizes ψ.

$$\mathcal{D}_2 \;=\; Tr[\,\varrho\;\log_2(\varrho)\,]\;=\;\log_2(\lambda_\varrho^{(1)})\,Tr\,[\,\varrho\;\,|e_1\rangle\,\langle e_1|\,]\;+\;\log_2(\lambda_\varrho^{(2)})\,Tr\,[\,\varrho\;\,|e_2\rangle\,\langle e_2|\,].$$

We use the Bloch sphere representation for ϱ in the expression for $\mathcal{D}_2$.

$$\psi \;=\; \frac{1}{2}\,(\,\mathcal{I}_2\;+\;\vec{\mathcal{V}}\bullet\vec{\sigma}\,).$$

$$\mathcal{D}_2 \;=\; Tr[\,\varrho\;\log_2(\psi)\,]\;=$$

$$\frac{1}{2}\;\log_2(\lambda_\psi^{(1)})\left[\,Tr[\,|e_1\rangle\,\langle e_1|\,]\;+\;\sum_i\;w_i\,Tr[\,\sigma_i\;\,|e_1\rangle\,\langle e_1|\,]\right]\;+$$

$$\frac{1}{2}\;\log_2(\lambda_\psi^{(2)})\left[\,Tr[\,|e_2\rangle\,\langle e_2|\,]\;+\;\sum_i\;w_i\,Tr[\,\sigma_i\;\,|e_2\rangle\,\langle e_2|\,]\right].$$

First note that $Tr[\,|e_1\rangle\,\langle e_1|\,]\;=\;Tr[\,|e_2\rangle\,\langle e_2|\,]\;=\;1$ since the $|e_j\rangle$ are projection operators.

Next define

$$\alpha_i^{(j)}\;=\;Tr[\sigma_i\;\,|e_j\rangle\,\langle e_j|]\;=\;\langle e_j|\;\sigma_i\;\,|e_j\rangle\,.$$

Evaluating these six (i = 1,2,3 and j = 1,2) constants yields:

$$\alpha_1^{(1)}\;=\;\frac{v_1\,\left(\sqrt{{v_2}^2+{v_3}^2+{v_1}^2}+v_3\right)}{{v_1}^2+\sqrt{{v_2}^2+{v_3}^2+{v_1}^2}v_3+{v_2}^2+{v_3}^2}\;=\;\frac{v_1\,(\,q\;+\;v_3\,)}{q^2\;+\;q\,v_3}\;=\;\frac{v_1}{q}.$$

$$\alpha_2^{(1)}\;=\;\frac{v_2\,\left(\sqrt{{v_2}^2+{v_3}^2+{v_1}^2}+v_3\right)}{{v_1}^2+\sqrt{{v_2}^2+{v_3}^2+{v_1}^2}v_3+{v_2}^2+{v_3}^2}\;=\;\frac{v_2\,(\,q\;+\;v_3\,)}{q^2\;+\;q\,v_3}\;=\;\frac{v_2}{q}.$$

$$\alpha_3^{(1)}\;=\;\frac{\left(\sqrt{{v_2}^2+{v_3}^2+{v_1}^2}+v_3\right)v_3}{{v_1}^2+\sqrt{{v_2}^2+{v_3}^2+{v_1}^2}v_3+{v_2}^2+{v_3}^2}\;=\;\frac{v_3\,(\,q\;+\;v_3\,)}{q^2\;+\;q\,v_3}\;=\;\frac{v_3}{q}.$$

$$\alpha_1^{(2)}\;=\;\frac{v_1\,\left(\sqrt{{v_2}^2+{v_3}^2+{v_1}^2}-v_3\right)}{-{v_1}^2+\sqrt{{v_2}^2+{v_3}^2+{v_1}^2}v_3-{v_2}^2-{v_3}^2}\;=\;-\frac{v_1\,(\,q\;-\;v_3\,)}{q^2\;-\;q\,v_3}\;=\;-\frac{v_1}{q}.$$

$$\alpha_2^{(2)}\;=\;\frac{v_2\,\left(\sqrt{{v_2}^2+{v_3}^2+{v_1}^2}-v_3\right)}{-{v_1}^2+\sqrt{{v_2}^2+{v_3}^2+{v_1}^2}v_3-{v_2}^2-{v_3}^2}\;=\;-\frac{v_1\,(\,q\;-\;v_3\,)}{q^2\;-\;q\,v_3}\;=\;-\frac{v_2}{q}.$$

$$\alpha_3^{(2)} \;=\; \frac{\left(\sqrt{{v_2}^2+{v_3}^2+{v_1}^2}-v_3\right)v_3}{-{v_1}^2+\sqrt{{v_2}^2+{v_3}^2+{v_1}^2}v_3-{v_2}^2-{v_3}^2} \;=\; -\frac{v_3\;(\;q\;-\;v_3\;)}{q^2\;-\;q\,v_3} \;=\; -\frac{v_3}{q}.$$

Putting it all together yields:

$$\mathcal{D}_2 \;=\; Tr[\;\varrho\;\log_2(\psi)\;]$$

$$=\;\frac{1}{2}\;\log_2\left(\lambda_\psi^{(1)}\right)\left[\;1\;+\;\sum_i\;w_i\,\alpha_i^{(1)}\;\right]\;+\;\frac{1}{2}\;\log_2\left(\lambda_\psi^{(2)}\right)\left[\;1\;+\;\sum_i\;w_i\,\alpha_i^{(2)}\;\right]$$

$$=\;\frac{1}{2}\left[\;1\;+\;\sum_i\;w_i\,\frac{v_i}{q}\;\right]\;\log_2\left(\lambda_\psi^{(1)}\right)\;+\;\frac{1}{2}\left[\;1\;+\;\sum_i\;w_i\,\frac{-\,v_i}{q}\;\right]\;\log_2\left(\lambda_\psi^{(2)}\right)$$

$$=\;\frac{1}{2}\left[\;1\;+\;\frac{\vec{\mathcal{W}}\bullet\vec{\mathcal{V}}}{q}\;\right]\;\log_2\left(\lambda_\psi^{(1)}\right)\;+\;\frac{1}{2}\left[\;1\;-\;\frac{\vec{\mathcal{W}}\bullet\vec{\mathcal{V}}}{q}\;\right]\;\log_2\left(\lambda_\psi^{(2)}\right).$$

Plugging in for the eigenvalues $\lambda_\psi^{(1)}$ and $\lambda_\psi^{(2)}$ which we found above yields:

$$\mathcal{D}_2 \;=\; Tr[\;\varrho\;\log_2(\psi)\;]$$

$$=\;\frac{1}{2}\left[\;1\;+\;\frac{\vec{\mathcal{W}}\bullet\vec{\mathcal{V}}}{q}\;\right]\;\log_2\left(\frac{1\;+\;q}{2}\right)\;+\;\frac{1}{2}\left[\;1\;-\;\frac{\vec{\mathcal{W}}\bullet\vec{\mathcal{V}}}{q}\;\right]\;\log_2\left(\frac{1\;-\;q}{2}\right)$$

$$=\;\frac{1}{2}\;\log_2(1\;-\;q^2\;)\;-\;1\;+\;\frac{\vec{\mathcal{W}}\bullet\vec{\mathcal{V}}}{2\,q}\;\log_2\left(\frac{1\;+\;q}{1\;-\;q}\right).$$

Putting all the pieces together to obtain $\mathcal{D}(\,\varrho\,\|\,\psi\,)$, we find

$$\mathcal{D}(\,\varrho\,\|\,\psi\,)\;=\;\mathcal{D}_1\;-\;\mathcal{D}_2$$

$$=\;\frac{1}{2}\log_2(1\;-\;r^2)\;+\;\frac{r}{2}\log_2\left(\frac{1\;+\;r}{1\;-\;r}\right)\;-\;\frac{1}{2}\;\log_2(1\;-\;q^2\;)\;-\;\frac{\vec{\mathcal{W}}\bullet\vec{\mathcal{V}}}{2\,q}\;\log_2\left(\frac{1\;+\;q}{1\;-\;q}\right)$$

$$=\;\frac{1}{2}\;\log_2(1\;-\;r^2)\;+\;\frac{r}{2}\;\log_2\left(\frac{1\;+\;r}{1\;-\;r}\right)\;-\;\frac{1}{2}\;\log_2(1\;-\;q^2\;)\;-\;\frac{r\;\cos(\theta)}{2}\;\log_2\left(\frac{1\;+\;q}{1\;-\;q}\right),$$

where θ is the angle between $\vec{\mathcal{W}}$ and $\vec{\mathcal{V}}$.

$\triangle$ - *End of Proof II.*

Ordinarily, $\mathcal{D}(\rho\|\phi)\;\neq\;\mathcal{D}(\phi\|\rho)$. However, when $r=q$, we can see from the above formula that $\mathcal{D}(\rho\|\phi)\;=\;\mathcal{D}(\phi\|\rho)$.

A few special cases of $\mathcal{D}(\rho\|\phi)$ are worth examining. Consider the case when $\phi = \frac{1}{2}\mathcal{I}_2$. In this case, $q = 0$, and

$$\mathcal{D}(\,\rho\,||\,\phi\,) \;=\; \frac{1}{2}\,\log_2\left(1-r^2\right) \;+\; \frac{r}{2}\,\log_2\left(\frac{1+r}{1-r}\right)$$

$$=\; \frac{1\;+\;r}{2}\,\log_2\left(\frac{1\;+\;r}{2}\right) \;+\; \frac{1\;+\;r}{2} \;+\; \frac{1\;-\;r}{2}\,\log_2\left(\frac{1\;-\;r}{2}\right) \;+\; \frac{1\;-\;r}{2} \;=\; 1\;-\;\mathcal{S}(\rho).$$

Thus $\mathcal{D}\left(\,\rho\,\left|\right|\,\frac{1}{2}\,\mathcal{I}_2\,\right) \;=\; 1\;-\;\mathcal{S}(\rho)$, where $\mathcal{S}(\rho)$ is the von Neumann entropy of ρ, the first density matrix in the relative entropy function. Note that in general

$$\mathcal{D}\left(\,\rho\,\left|\right|\,\frac{1}{d}\,\mathcal{I}_d\,\right) \;=\; \log_2(d)\;-\;\mathcal{S}(\rho)$$

since

$$\mathcal{D}\left(\,\rho\,\left|\left|\right|\,\frac{1}{d}\,\mathcal{I}_d\,\right) \;=\; Tr\left[\,\rho\,\left(\,\log_2(\rho)\;-\;\log_2\left(\frac{1}{d}\,\mathcal{I}_d\right)\,\right)\,\right]$$

$$=\; Tr\left[\,\rho\,\left(\,\log_2(\rho)\;+\;\log_2(d)\,\mathcal{I}_d\,\right)\,\right] \;=\; \log_2(d)\,Tr\left[\,\rho\,\right]\;-\;\mathcal{S}(\rho) \;=\; \log_2(d)\;-\;\mathcal{S}(\rho).$$

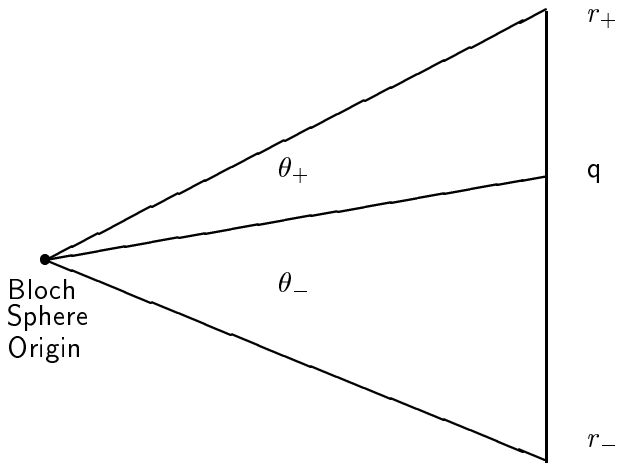

Figure 2.25: Definition of the Bloch vectors $\vec{r}_+$, $\vec{q}$, and $\vec{r}_-$ used in the derivation below.

XI Appendix B: The Linear Channel Transcendental Equation

In this appendix, we derive the transcendental equation for determining the optimum position of the average density matrix for a linear channel. The picture of the quantities we shall define shortly is below.

We assume that in general all the $\{ t_k \neq 0\}$. We also assume the linear channel is oriented in the z direction, so that $\lambda_x = \lambda_y = 0$, but $\lambda_z \neq 0$. We define

$$A = t_x^2 + t_y^2 + (t_z + \lambda_z)^2 = r_+^2.$$

$$B = t_x^2 + t_y^2 + (t_z + \beta \lambda_z)^2 = q(\beta)^2.$$

$$C = t_x^2 + t_y^2 + (t_z - \lambda_z)^2 = r_-^2.$$

The three quantities above refer respectively to the distance from the Bloch sphere origin to r_+, the optimum point q we seek, and r_-. We define the three Bloch vectors $\vec{r}_+$, $\vec{q}$ and $\vec{r}_-$ in Figure 2.25 above, and refer to their respective magnitudes as r_+, q, and r_-. Here $\beta \in [-1, 1]$, so that q can range along the entire line segment between r_+ and r_-.

As discussed in the Linear Channels section of this chapter, the condition on q is that $\mathcal{D}(\, r_+ \,\|\, q\,) \;=\; \mathcal{D}(\, r_- \,\|\, q\,)$.

Recall that

$$\mathcal{D}(\, r \,\|\, q\,) \;=\; =\; \frac{1}{2}\,\log_2(1-r^2) + \frac{r}{2}\,\log_2\left(\frac{1\;+\;r}{1\;-\;r}\right) - \frac{1}{2}\;\log_2(1-q^2\;) - \frac{r\;\cos(\theta)}{2}\;\log_2\left(\frac{1\;+\;q}{1\;-\;q}\right),$$

where θ is the angle between r and q. To determine θ, we use the law of cosines. If θ is the angle between sides a and b of a triangle with sides a, b and c, then we have:

$$\cos(\theta) \;=\; \frac{a^2\;+\;b^2\;-\;c^2}{2\,a\,b}.$$

Our condition $\mathcal{D}(\, r_+ \,\|\, q\,) \;=\; \mathcal{D}(\, r_- \,\|\, q\,)$ becomes:

$$\frac{1}{2}\,\log(1\;-\;r_+^2)\;+\;\frac{r}{2}\,\log\left(\frac{1\;+\;r_+}{1\;-\;r_+}\right)\;-\;\frac{r_+\;\cos(\theta_+)}{2}\;\log\left(\frac{1\;+\;q}{1\;-\;q}\right)$$

$$=\;\frac{1}{2}\,\log(1\;-\;r_-^2)\;+\;\frac{r}{2}\,\log\left(\frac{1\;+\;r_-}{1\;-\;r_-}\right)\;-\;\frac{r_-\;\cos(\theta_-)}{2}\;\log\left(\frac{1\;+\;q}{1\;-\;q}\right),$$

where we canceled the term which was identically a function of q from both sides, and converted all logs from base 2 to natural logs by multiplying both sides by log(2).

Determining θ_+ and θ_-, we find:

$$\cos(\theta_+) \;=\; \frac{r_+^2\;+\;q^2\;-\;(\,(\,1\;-\;\beta\,)\,\lambda_z\,)^2}{2\,q\,r_+}.$$

$$\cos(\theta_-) \;=\; \frac{r_-^2\;+\;q^2\;-\;(\,(\,1\;+\;\beta\,)\,\lambda_z\,)^2}{2\,q\,r_-}.$$

Next, recall the identity

$$\tanh^{(-1)}[\,x\,] \;=\; \frac{1}{2}\,\log\left(\frac{1\;+\;x}{1\;-\;x}\right).$$

Using this identity for arctanh, our relative entropy equality relation between the two end-

points of the linear channel becomes:

$$\frac{1}{2}\log(1\,-\,A)\,+\,\sqrt{A}\,\tanh^{(-1)}\left(\sqrt{A}\right)\,-\,\frac{\sqrt{A}\,(\,A\,+\,B\,-\,((1\,-\,\beta)\,\lambda_z)^2\,)}{2\,\sqrt{A\,B}}\,\tanh^{(-1)}\left(\sqrt{B}\right)$$

$$=\,\frac{1}{2}\,\log(1-C)+\sqrt{C}\,\tanh^{(-1)}\left(\sqrt{C}\right)-\frac{\sqrt{C}\,(\,C\,+\,B\,-\,((1\,+\,\beta)\,\lambda_z)^2\,)}{2\sqrt{B\,C}}\,\tanh^{(-1)}\left(\sqrt{B}\right).$$

We can cancel several terms to obtain

$$\frac{1}{2}\,\log(1\,-\,A)\,+\,\sqrt{A}\,\tanh^{(-1)}\left(\sqrt{A}\right)\,-\,\frac{(\,A\,+\,2\,\beta\,\lambda_z^2\,)}{2\,\sqrt{B}}\,\tanh^{(-1)}\left(\sqrt{B}\right)$$

$$=\,\frac{1}{2}\,\log(1\,-\,C)\,+\,\sqrt{C}\,\tanh^{(-1)}\left(\sqrt{C}\right)\,-\,\frac{(\,C\,-\,2\,\beta\,\lambda_z^2\,)}{2\,\sqrt{B}}\,\tanh^{(-1)}\left(\sqrt{B}\right),$$

which in turn becomes:

$$\frac{1}{2}\,\log(1\,-\,A)\,+\,\sqrt{A}\,\tanh^{(-1)}\left(\sqrt{A}\right)\,-\,\frac{(\,A\,-\,C\,+\,4\,\beta\,\lambda_z^2\,)}{2\,\sqrt{B}}\,\tanh^{(-1)}\left(\sqrt{B}\right)$$

$$=\,\frac{1}{2}\,\log(1\,-\,C)\,+\,\sqrt{C}\,\tanh^{(-1)}\left(\sqrt{C}\right).$$

Using our definitions above for A and C, we find that A - C = $4\,\lambda_z\,t_z$. Substituting this into the relation immediately above yields:

$$\frac{1}{2}\,\log(1\,-\,A)\,+\,\sqrt{A}\,\tanh^{(-1)}\left(\sqrt{A}\right)\,-\,\frac{(\,4\,\lambda_z\,t_z\,+\,4\,\beta\,\lambda_z^2\,)}{2\,\sqrt{B}}\,\tanh^{(-1)}\left(\sqrt{B}\right)$$

$$=\,\frac{1}{2}\log(1\,-\,C)\,+\,\sqrt{C}\,\tanh^{(-1)}\left(\sqrt{C}\right),$$

which we adjust to our final answer:

$$\frac{4\,\lambda_z\,(\,t_z\,+\,\beta\,\lambda_z\,)}{\sqrt{B}}\,\tanh^{(-1)}\left(\sqrt{B}\right)$$

$$=\,\log(1\,-\,A)\,-\,\log(1\,-\,C)\,+\,2\,\sqrt{A}\,\tanh^{(-1)}\left(\sqrt{A}\right)\,-\,2\,\sqrt{C}\,\tanh^{(-1)}\left(\sqrt{C}\right).$$

Note that B is a function of β, so the entire functionality of β lies to the left of the equality sign in the expression above. All terms on the right hand side are functions of the $\{t_k\}$

and $\{\lambda_z\}$, so the right hand side is a constant while we vary β. Since all the functions of β on the left hand side are smooth functions, the search for the optimum $\beta \equiv q$, although transcendental, is well behaved and fairly easy.

XII Appendix C: Quantum Channel Descriptions

The Kraus quantum channel representation is given by the set of Kraus matrices $\mathcal{A} = \{ A_i \}$ which represent the channel dynamics via the relation:

$$\mathcal{E}(\rho) = \sum_i A_i \, \rho \, A_i^\dagger .$$

The normalization requirement for the Kraus matrices is:

$$\sum_i A_i^\dagger \, \rho \, A_i = I.$$

A channel is unital if it maps the identity to the identity. This requirement becomes, upon setting $\rho = \mathrm{I}$:

$$\sum_i A_i \, \rho \, A_i^\dagger = \sum_i A_i \, A_i^\dagger = I.$$

For qubit channels, the set of Kraus operators, $\{ \mathcal{A}_i \}$ can mapped to a set of King-Ruskai-Szarek-Werner ellipsoid channel parameters $\{ t_k \, , \, \lambda_k \}$, where $k = 1, 2, 3$.

XII.a The Two Pauli Channel Kraus Representation

$$A_1 = \begin{bmatrix} \sqrt{x} & 0 \\ 0 & \sqrt{x} \end{bmatrix} \qquad A_2 = \sqrt{\frac{1 - x}{2}} \; \sigma_x = \begin{bmatrix} 0 & \sqrt{\frac{1-x}{2}} \\ \sqrt{\frac{1-x}{2}} & 0 \end{bmatrix}.$$

$$A_3 = - i \sqrt{\frac{1 - x}{2}} \quad \sigma_y = \begin{bmatrix} 0 & - \sqrt{\frac{1-x}{2}} \\ \sqrt{\frac{1-x}{2}} & 0 \end{bmatrix}.$$

In words, the channel leaves the qubit transiting the channel alone with probability x, and does a σ_x on the qubit with probability $\frac{1-x}{2}$ or does a σ_y on the qubit with probability $\frac{1-x}{2}$. The Two Pauli channel is a unital channel. The corresponding King-Ruskai-Szarek-Werner ellipsoid channel parameters are $t_x = t_y = t_z = 0$, and $\lambda_x = \lambda_y = x$, while $\lambda_z = 2\,x - 1$ [25]. Here $x \in [\,0\,,\,1\,]$.

XII.b The Depolarization Channel Kraus Representation

$$A_1 = \begin{bmatrix} \sqrt{x} & 0 \\ 0 & \sqrt{x} \end{bmatrix} \qquad A_2 = \sqrt{\frac{1-x}{3}} \quad \sigma_x = \begin{bmatrix} 0 & \sqrt{\frac{1-x}{3}} \\ \sqrt{\frac{1-x}{3}} & 0 \end{bmatrix}.$$

$$A_3 = -i\sqrt{\frac{1-x}{3}} \quad \sigma_y = \begin{bmatrix} 0 & -\sqrt{\frac{1-x}{3}} \\ \sqrt{\frac{1-x}{3}} & 0 \end{bmatrix}.$$

$$A_4 = \sqrt{\frac{1-x}{3}} \quad \sigma_z = \begin{bmatrix} \sqrt{\frac{1-x}{3}} & 0 \\ 0 & -\sqrt{\frac{1-x}{3}} \end{bmatrix}.$$

In words, the channel leaves the qubit transiting the channel alone with probability x, and does a σ_x on the qubit with probability $\frac{1-x}{3}$ or does a σ_y on the qubit with probability $\frac{1-x}{3}$. or does a σ_z on the qubit with probability $\frac{1-x}{3}$. The Depolarization channel is a unital channel. The corresponding King-Ruskai-Szarek-Werner ellipsoid channel parameters are $t_x = t_y = t_z = 0$, and $\lambda_x = \lambda_y = \lambda_z = \frac{4x-1}{3}$ [25]. Here $x \in [0, 1]$.

XII.c The Amplitude Damping Channel Kraus Representation

$$A_1 = \begin{bmatrix} \sqrt{x} & 0 \\ 0 & 1 \end{bmatrix} \qquad A_2 = \begin{bmatrix} 0 & 0 \\ \sqrt{1-x} & 0 \end{bmatrix}.$$

In this scenario, the channel leaves untouched a spin down qubit. For a spin up qubit, with probability x it leaves the qubit alone, while with probability 1 - x the channel flips the spin from up to down. Thus, when $x = 0$, every qubit emerging from the channel is in the spin down state. The Amplitude Damping channel is *not* a unital channel. The corresponding King-Ruskai-Szarek-Werner ellipsoid channel parameters are $t_x = 0$, $t_y = 0$, $t_z = 1 - x$, $\lambda_x = \sqrt{x}$, $\lambda_y = \sqrt{x}$, and $\lambda_z = x$ [25]. Here $x \in [0, 1]$.

XIII Appendix D: Numerical Analysis of Optimal Signal Ensembles using MAPLE and MATLAB

The iterative, relative entropy based algorithm outlined above was implemented in MAPLE, and provided the plots and numbers cited in this chapter. In addition, numerical answers were verified using a brute force algorithm based on MATLAB's Optimization Toolbox. The MATLAB optimization criterion was the channel output Holevo χ quantity. Input qubit ensembles of two, three and four states were used. After channel evolution, the output ensemble Holevo χ was calculated. With this function specified as to be maximized, the MATLAB Toolbox varied the parameters for the ensemble qubit input pure states and the states corresponding a priori probabilities. Pure state qubits were represented as:

$$|\,\psi\,\rangle \;=\; \left[\begin{array}{c} \alpha \\ \sqrt{1\,-\,\alpha^2}\;e^{i\,\theta} \end{array}\right],$$

thereby requiring two parameters, $\{\,\alpha\,,\,\theta\,\}$, for each input qubit state. Thus a two state input qubit ensemble required an optimization over a space of dimension five, when the a priori probabilities are included. Three and four state ensembles required optimization over spaces of dimension eight and eleven, respectively.

Chapter 3

Numerical Explorations of Channel Additivity for Qubit Quantum Channels

I Abstract

In this chapter, we consider the issue of channel additivity for the transmission of classical information through quantum channels. The Holevo-Schumacher-Westmoreland (HSW) (product state) channel capacity $\mathcal{C}_1$ and probability of error (P_e) criteria are considered[2]. We examine numerically $\mathcal{C}_1$ additivity for three different qubit channels: The Depolarization, Two Pauli and Amplitude Damping channels. Our results confirm analytical predictions made by Christopher King[19] in regards to the additivity of qubit unital channels, and indicate similar additivity considerations extend to non-unital qubit channels such as the Amplitude Damping channel. In addition, we examine whether entanglement across input signalling states for parallel combinations of channels reduces the overall signalling probability of error (P_e).

I.a Brief Review of Channel Additivity

As discussed in Sections II, III, and IV in Chapter 1, the general channel picture can be summarized as shown in Figure 3.1.

Recall the discussion of channel additivity from Section V of Chapter 1. Our model of multiqubit channels is that the channel acts *independently* on each qubit passing through

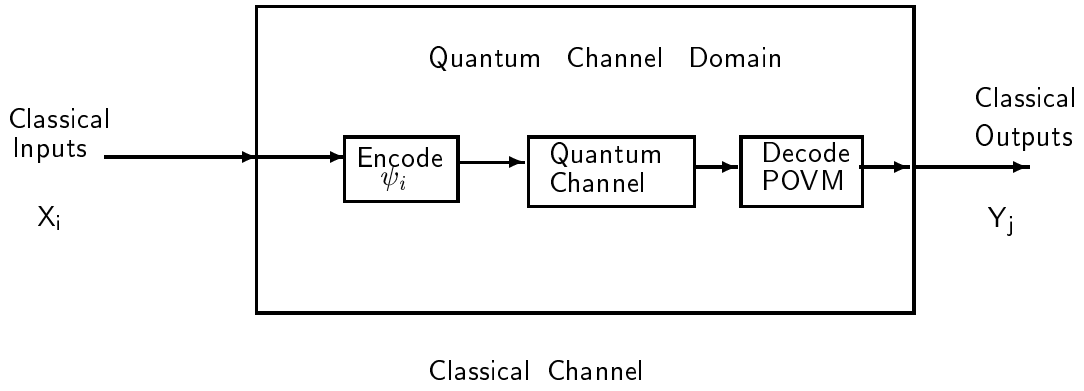

Figure 3.1: Sending classical information over a quantum channel.

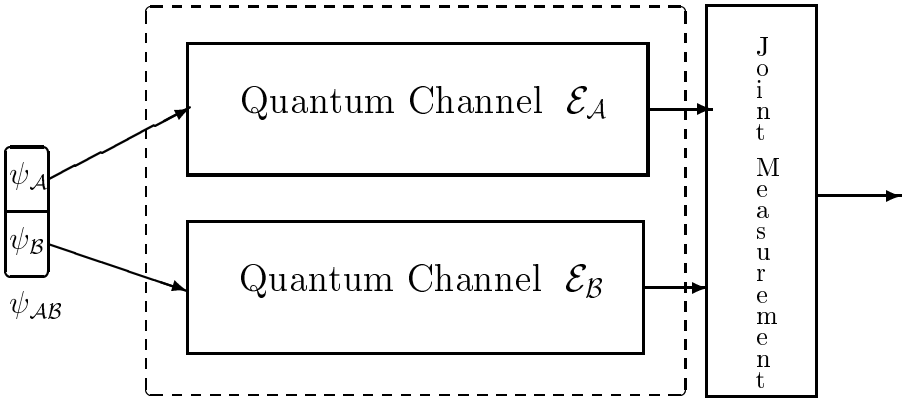

Figure 3.2: Tensor product of two channels.

the channel. Such a model does not allow correlated errors to act on the qubits passing through the channel. For channel additivity, we are interested in the scenario where the single quantum channel in Figure 3.1 is composed of the tensor product of two subchannels, as shown in Figure 3.2[1].

The dotted box in Figure 3.2 is meant to represent the single quantum channel shown in Figure 3.1. In this chapter, the "subchannels" $\mathcal{E}_{\mathcal{A}}$ and $\mathcal{E}_{\mathcal{B}}$ will always be single qubit channels.

[1]Please see Appendix A for a mathematical description of tensor product channels.

There are two common criteria for measuring the quality of the transmission of classical information over a channel, regardless of whether the channel is classical or quantum. These criteria are the (Product State) Channel Capacity $\mathcal{C}_1$ [12, 13, 14] and probability of error (P_e)[38]. We are interested in the additivity of $\mathcal{C}_1$, and for the probability of error (P_e) criterion, whether entanglement across input signalling states reduces the P_e over signalling ensembles consisting of product states (i.e.: no entanglement). In other words, does allowing entangled inputs ψ_{AB} and joint measurements, as shown in Figure 3.2, improve $\mathcal{C}_1$ and P_e performance over the use of product state inputs and independent measurements at each "sub-channel" output?

It should be noted King and Ruskai [39] have shown that conducting independent measurements at the subchannel outputs destroys any $\mathcal{C}_1$ capacity benefits arising from the use of entangled inputs. Thus, if we would like to use entangled input signalling states, we must use a joint measurement scheme across the sub-channel outputs.

It is interesting that the reverse does not hold. That is, using input product state signals and joint measurements, one can, for certain non-unital channels, achieve improved channel capacity performance over the use of input product states signals and independent subchannel measurements[39].

I.b Two Pauli Channels

Our interest in the topic of additivity was spurred by the work of Bennett, Fuchs and Smolin[38]. In that paper, the Two Pauli channel was studied. (See Appendix A for a more complete description of the Two Pauli Channel.) The authors used probability of error (P_e) as a channel criteria, and found that there was a *slight* improvement in using generic two qubit states (i.e.: allowing entanglement) as signaling states versus two qubit product states when the channel parameter x exceeded $\frac{1}{3}$. That is, entanglement between input signalling states improved the P_e performance over two independent uses of the Two Pauli qubit channel.

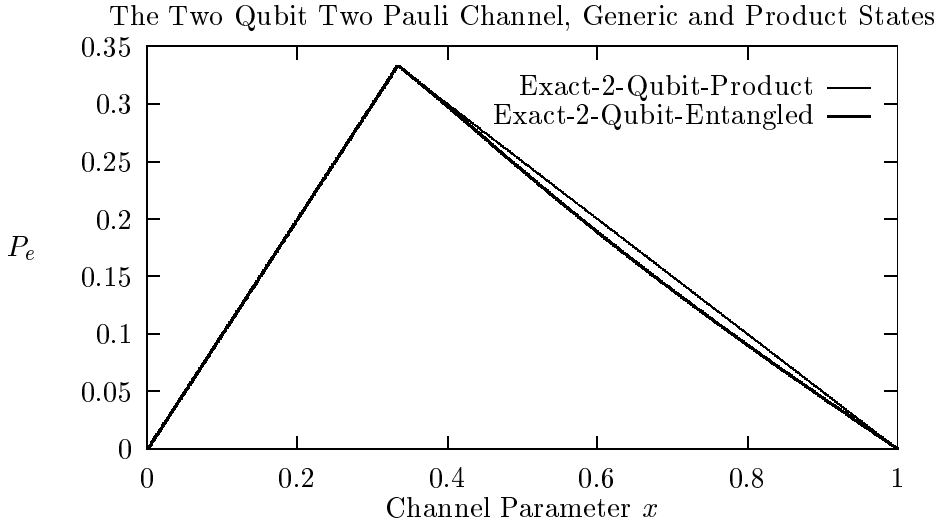

Figure 3.3: P_e for the tensor product of wo Two Pauli channels.

Unfortunately, the work of Christopher King indicates that for two uses of the Two Pauli channel, the entanglement enhanced behavior of the probability of error *does not* correspond to superadditivity of $\mathcal{C}_1$.

I.c Numerical Results - The Two Pauli Channel

The plot in Figure 3.3 is an analytic expression derived in the Bennett-Fuchs-Smolin paper[38]. In the plots to follow, we will be linearly interpolating between numerically determined data points. The data points will be indicated by either bullets or open circles. The bullet and circle centers indicate the data points in question.

A maximum improvement in probability of error (P_e) for the two qubit Two Pauli channel of 0.0124, or 93 percent of the product state P_e, occurred at $x = 0.6631$. Intuitively, one would have thought there should be more separation between the product state and generic state P_e curves, given that in the two qubit scenario, there are an additional two degrees of freedom (i.e.: 6 dimensions versus 4 dimensions) available to the generic states than the

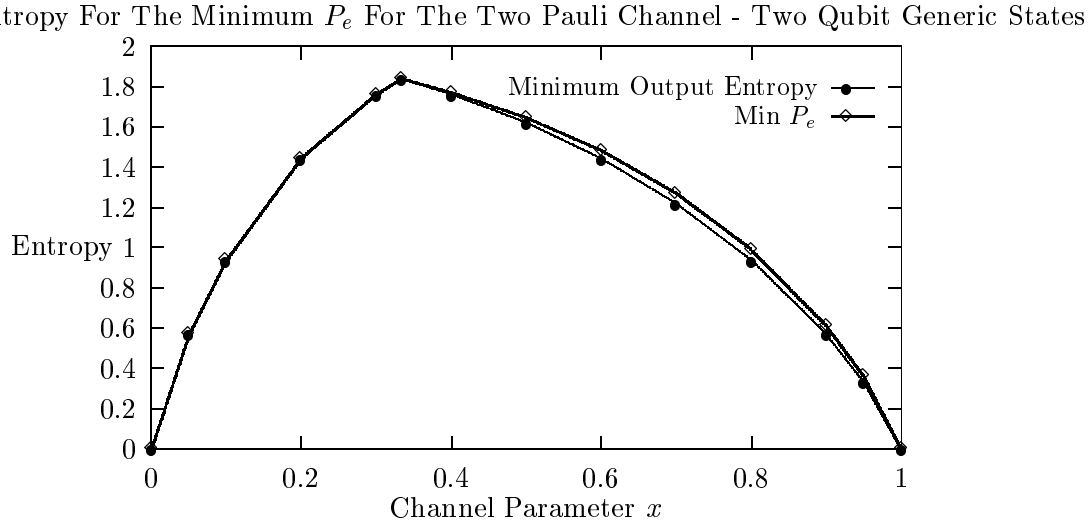

Figure 3.4: Entropy of the optimum P_e output signalling states versus the minimum output entropy (i.e.: the optimum $\mathcal{C}_1$ signalling) states.

product state signals[2]. The P_e optimum product states had the minimum possible output entropy, while the generic input states, allowing entanglement between the two qubits, had slightly greater than the minimum output entropy for $x \in [\frac{1}{3}, 1]$. For $x \in [0, \frac{1}{3}]$, all minimum P_e solutions were product states.

It is interesting to examine how much entanglement the optimal P_e signalling states use when generic inputs are allowed. As shown in Figure 3.4, the answer is very little.

The bullet curve in Figure 3.4 is the minimum possible entropy at the (two qubit) channel output. The diamond curve is the channel output entropy for the optimal P_e signalling states. One can conclude the P_e optimization process uses only a minimal amount of entanglement above and beyond what it has too. This topic will be discussed further in the chapter summary.

[2]See Section V of Chapter 1 for how the degrees of freedom of the various qubit states are calculated.

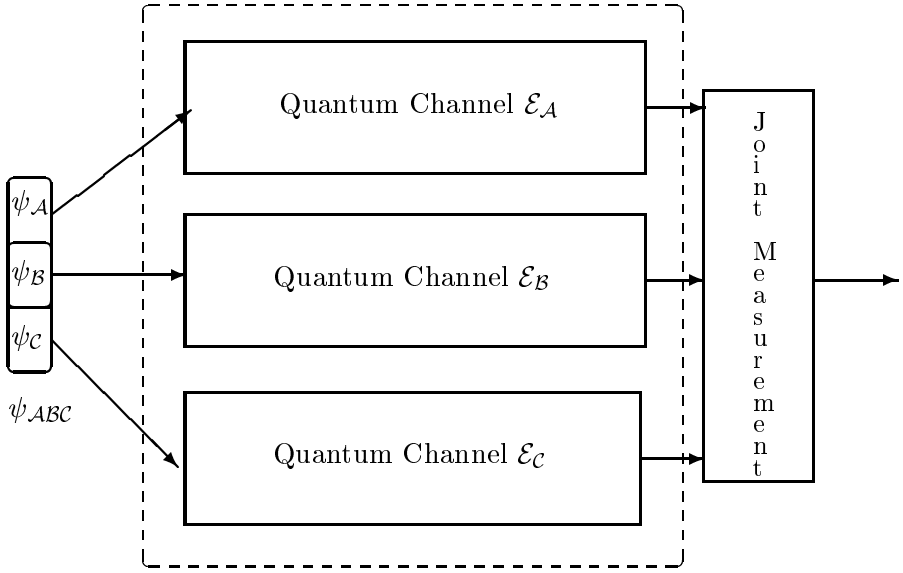

Figure 3.5: Tensor product of three channels.

I.d Three Two Pauli Channels in Parallel

To investigate further, an analysis of three Two Pauli channels acting in parallel was conducted.

Here $\mathcal{E}_\mathcal{A}$, $\mathcal{E}_\mathcal{B}$, and $\mathcal{E}_\mathcal{C}$ are single qubit Two Pauli channels. In this case, there are six free parameters (degrees of freedom) for the 3 qubit product states, and 14 free parameters (degrees of freedom) for the generic 3 qubit states. Numerical simulations indicated there is even less of a difference for the probability of error (P_e) between the generic and product state input scenarios in the three qubit case as compared to the two qubit case.

There was no difference between input and generic state input P_e performance seen over the channel parameter range $x \in [0, \frac{1}{3}]$, just as in the two qubit case. For $x \in [\frac{1}{3}, 1]$, the difference between product and generic state inputs P_e performance was of $O\left(10^{-6}\right)$, which was the noise floor for the numerics. Thus, there was much less of a difference in P_e performance in the three qubit case than that seen in the two qubit case. At least for the Two Pauli channel, the slight benefit of using generic states over product states essentially

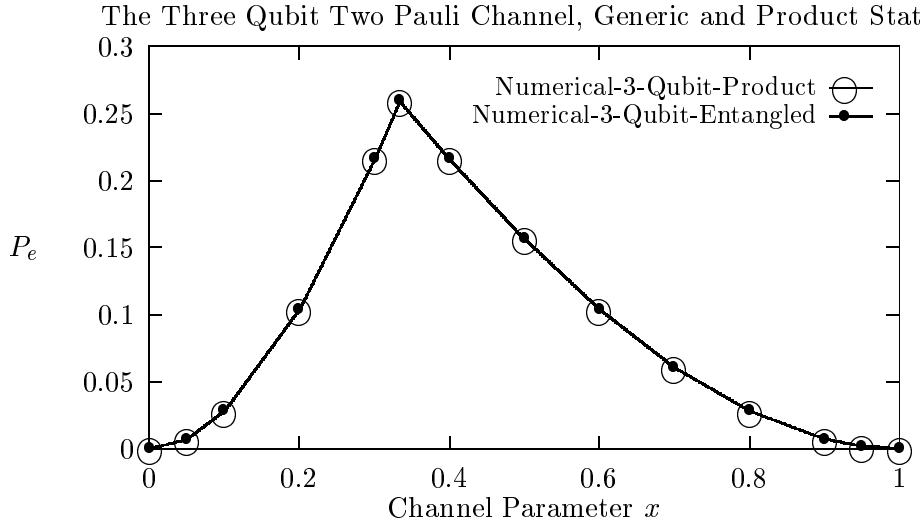

Figure 3.6: P_e for three tensored Two Pauli channels.

vanished when we moved from two qubit to three qubit joint signaling states.

Now let us look at the additivity of $\mathcal{C}_1$ for two Two Pauli channel's in parallel, as shown in Figure 3.2. From the work of Christopher King[19], and the fact that the Two Pauli channel is a unital channel, we know:

1) C_1 is additive, meaning $\mathcal{C}_1^{\mathcal{A}\otimes\mathcal{B}} = \mathcal{C}_1^{\mathcal{A}} + \mathcal{C}_1^{\mathcal{B}}$.

The KRSW qubit channel formalism implies all qubit unital channels are diagonal in some operator basis. Therefore, our analytical work in Chapter 1 leads to the following additional facts.

2) The average signalling state should be[3] I_4.

3) The optimum input signaling states for the tensor channel $\mathcal{A}\otimes\mathcal{B}$ are a subset of the input states which yield the minimum entropy at the output of $\mathcal{E}^{\mathcal{A}\otimes\mathcal{B}}$. Similarly, the optimum input signaling states for the $\mathcal{A}$ channel are a subset of the input states which yield the

[3]See Sections X.c and XI of Chapter 1.

minimum entropy at the output of $\mathcal{E}^{\mathcal{A}}$, and the optimum input signaling states for the $\mathcal{B}$ channel are a subset of the input states which yield the minimum entropy at the output of $\mathcal{E}^{\mathcal{B}}$. In equation's these relations are:

$$\mathcal{C}_1^{\mathcal{A}\otimes\mathcal{B}} \;=\; \log_2(4) \;-\; min_{\rho_{\mathcal{AB}}}\, \mathcal{S}\left(\mathcal{E}(\rho_{\mathcal{AB}})\right).$$

$$\mathcal{C}_1^{\mathcal{A}} \;=\; \log_2(2) \;-\; min_{\rho_{\mathcal{A}}}\, \mathcal{S}\left(\mathcal{E}(\rho_{\mathcal{A}})\right).$$

$$\mathcal{C}_1^{\mathcal{B}} \;=\; \log_2(2) \;-\; min_{\rho_{\mathcal{B}}}\, \mathcal{S}\left(\mathcal{E}(\rho_{\mathcal{B}})\right).$$

$$\mathcal{C}_1^{\mathcal{A}\otimes\mathcal{B}} \;=\; \mathcal{C}_1^{\mathcal{A}} \;+\; \mathcal{C}_1^{\mathcal{B}}.$$

The $\mathcal{C}_1$ channel additivity of the two parallel Two Pauli channels therefore implies

4) the minimum output entropy of the channel is additive[4], meaning

$$min_{\rho_{\mathcal{AB}}}\, \mathcal{S}\left(\mathcal{E}(\rho_{\mathcal{AB}})\right) \;=\; min_{\rho_{\mathcal{A}}}\, \mathcal{S}\left(\mathcal{E}(\rho_{\mathcal{A}})\right) \;+\; min_{\rho_{\mathcal{B}}}\, \mathcal{S}\left(\mathcal{E}(\rho_{\mathcal{B}})\right).$$

The fact that the joint minimum output entropy states of $\mathcal{E}^{\mathcal{A}\otimes\mathcal{B}}$ are product states implies that $\mathcal{C}_1^{\mathcal{A}\otimes\mathcal{B}}$ can be achieved using an input ensemble consisting of product states. Note that the possibility exists that an input ensemble consisting of one or more entangled input states could *also* achieve $\mathcal{C}_1^{\mathcal{A}\otimes\mathcal{B}}$. Some types of channels implement *entanglement-breaking*, mapping entangled input states into unentangled (product) channel output states[41, 42]. However, it is interesting to note that our numerical simulations always converged to input ensembles consisting entirely of product signaling states, and never to an ensemble containing even a single entangled input state.

Our numerics confirmed these properties. Specifically, we found:

1) The $\mathcal{C}_1$ capacity is additive.

2) The optimal $\mathcal{C}_1$ input ensemble signalling states corresponded to minimum channel output

[4] See Sections XI and XIII of Chapter 1.

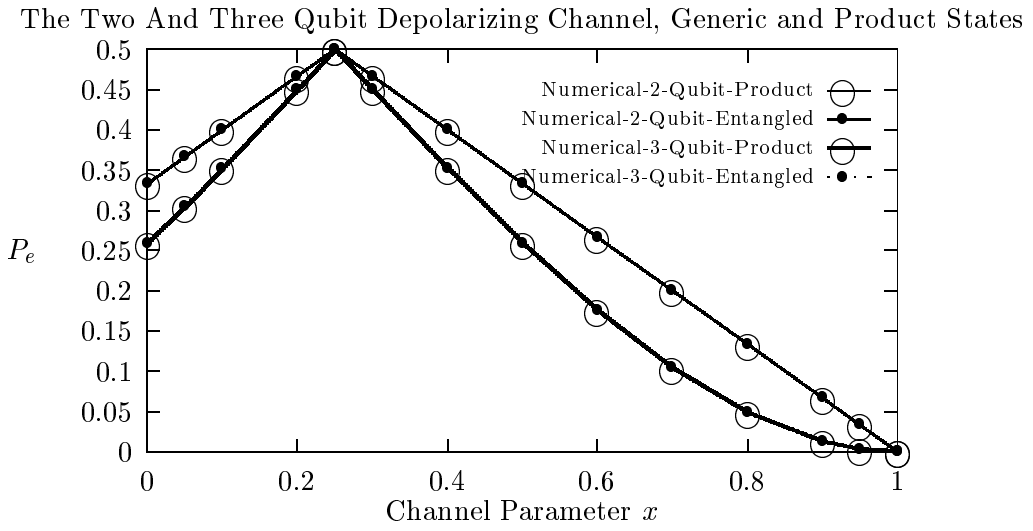

Figure 3.7: P_e for two and three parallel Depolarization channels.

entropy states.

3) The optimal $\mathcal{C}_1$ input signalling states were product states.

I.e Numerical Results - The Depolarization Channel

For the Depolarizing Channel, the optimum P_e signalling states for both the two and three qubit cases were product states.

In regards to the HSW capacity, the Depolarizing channel is a unital channel, and the same three $\mathcal{C}_1$ properties outlined for the Two Pauli channel apply to the Depolarizing channel. Our numerics confirmed these properties. Specifically, we found that:

1) The $\mathcal{C}_1$ capacity is additive.

2) The optimal $\mathcal{C}_1$ input ensemble signalling states corresponded to minimum channel output entropy states.

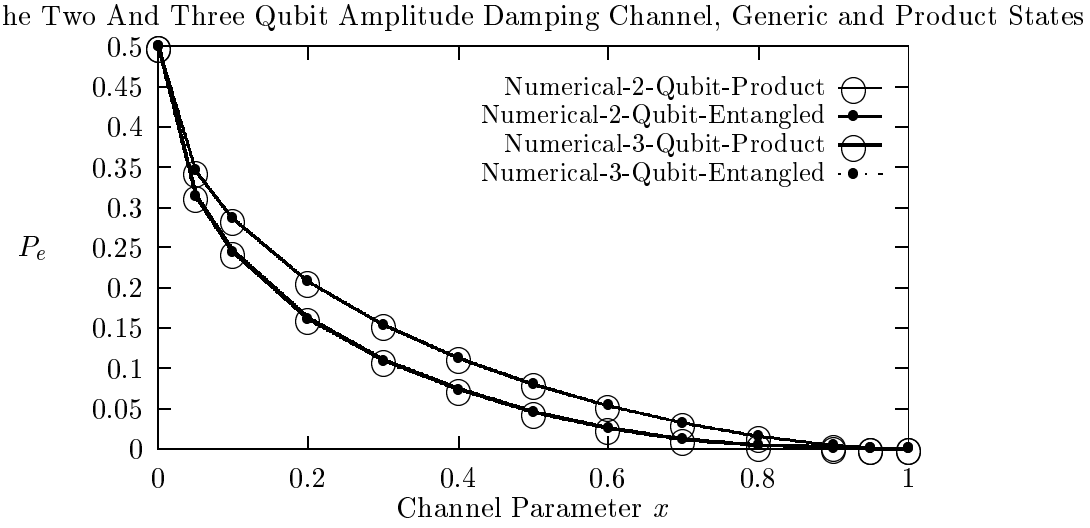

Figure 3.8: P_e for two and three parallel Amplitude Damping channels.

3) The optimum $\mathcal{C}_1$ input signalling states were product states.

I.f Numerical Results - The Amplitude Damping Channel

The Amplitude Damping channel is the only *non*-unital channel we numerically investigated.

For the Amplitude Damping channel, there was marginal improvement on the order of 10^{-3} for two qubit entangled states and $\approx 10^{-4}$ for three qubit states, in using generic input signalling states over product input signalling states for the P_e criterion.

The Amplitude Damping channel yielded results that differed in a number of respects from those for the Two Pauli and Depolarization channels. Recall that the Amplitude Damping channel is *not* unital. The analysis of Chapter 1 does not apply to non-unital channels, nor does the work of King and Ruskai[19, 25]. Yet the channel capacity $\mathcal{C}_1$ was numerically

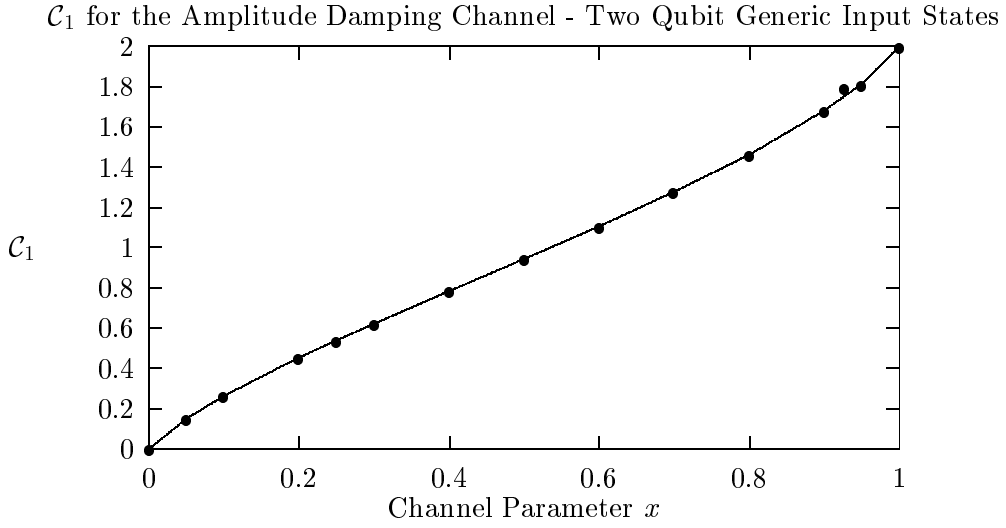

Figure 3.9: The $\mathcal{C}_1$ capacity for two parallel Amplitude Damping channels.

found to be strictly additive over two uses of the channels:

$$\mathcal{C}_1(Two\ Uses) \;=\; 2\,\mathcal{C}_1(One\ Use) \qquad \forall\, x \in [0,1].$$

Note that when $x = 0$, all input states are mapped to the same output state, resulting in a channel capacity of zero. At the other extreme, when $x = 1$, the channel is the identity channel on two qubits, and we expect, from the Holevo Theorem on Accessible Information[2], that the channel capacity should be two.

For the Amplitude Damping channel, the minimum output entropy is always zero, since the two qubit pure input product state consisting of spin down-down ($\downarrow_A \otimes \downarrow_B$) always passes through the channel untouched, and hence has output entropy equal to zero. Furthermore, this input state is the *unique* minimum entropy state. Therefore, the relation found for diagonal unital channels,

$$\mathcal{C}_1^{\mathcal{A}\otimes\mathcal{B}} \;=\; \log_2(4) \;-\; min_{\rho_{\mathcal{AB}}}\, \mathcal{S}\left(\mathcal{E}(\rho_{\mathcal{AB}})\right)$$

does *not* apply. Interestingly, the minimum output entropy of $\mathcal{E}^{\mathcal{A}\otimes\mathcal{B}}$ *is* additive. This is due to the fact that the minimum output entropy of $\mathcal{E}^{\mathcal{A}}$ is zero, as is the minimum output entropy of $\mathcal{E}^{\mathcal{B}}$, both of which are achieved using the spin down input state. The minimum output entropy of $\mathcal{E}^{\mathcal{A}\otimes\mathcal{B}}$ is also zero, achieved by using the pure input product state of two spin down qubits ($\rho_{AB} \sim \downarrow_A \otimes \downarrow_B$), which leads to

$$min_{\rho_{\mathcal{AB}}}\, \mathcal{S}\left(\mathcal{E}(\rho_{\mathcal{AB}})\right) \;=\; min_{\rho_{\mathcal{A}}}\, \mathcal{S}\left(\mathcal{E}(\rho_{\mathcal{A}})\right) \;+\; min_{\rho_{\mathcal{B}}}\, \mathcal{S}\left(\mathcal{E}(\rho_{\mathcal{B}})\right).$$

Peter Shor has shown a relation between the additivity of the minimum output entropy of the tensor product of two quantum channels, and the additivity of $\mathcal{C}_1$, namely additivity of the minimum output entropy implies additivity of $\mathcal{C}_1$, and vice versa[18]. While we shall not discuss this relation in detail, the minimum output entropy additivity relation for the tensor product combination of two Amplitude Damping channels agrees with the Shor result.

In the Amplitude Damping channel case, there is a unique minimum output entropy state, and thus the optimal signalling ensemble must contain output states which do not have the minimum output entropy, in contrast to the diagonal unital channel case. Furthermore, the input signaling states used in input ensembles achieving $\mathcal{C}_1$ were always found to be product states, although of course these states did not possess the minimum output entropy, since only one input state corresponds to the minimum output entropy state.

II Summary

In this chapter, our numerical simulations of the Two Pauli, Depolarization and Amplitude Damping qubit channels support the following conjectures:

1) The strict additivity of the Holevo-Schumacher-Westmoreland channel capacity for the tensor product of two (single qubit) quantum channels. Our support for this conjecture lies in the fact that for all three channels studied, we found $\mathcal{C}_1^{A\otimes B} \;=\; \mathcal{C}_1^{A} \;+\; \mathcal{C}_1^{B}$.

2) The Shor relation: The minimum channel output entropy of the tensor product of two quantum channels is additive if and only if the Holevo-Schumacher-Westmoreland channel capacity is strictly additive. Our support for this conjecture lies in the fact that for all three channels studied, we found

$$min_{\rho_{\mathcal{AB}}} \mathcal{S}\left(\mathcal{E}^{\mathcal{A}\otimes\mathcal{B}}(\rho_{\mathcal{AB}})\right) \;=\; min_{\rho_{\mathcal{A}}} \mathcal{S}\left(\mathcal{E}^{\mathcal{A}}(\rho_{\mathcal{A}})\right) \;+\; min_{\rho_{\mathcal{B}}} \mathcal{S}\left(\mathcal{E}^{\mathcal{B}}(\rho_{\mathcal{B}})\right)$$

and

$$\mathcal{C}_1^{A\otimes B} \;=\; \mathcal{C}_1^{A} \;+\; \mathcal{C}_1^{B}.$$

For the probability of error criterion, entanglement across the input states of parallel, independent quantum channels provides either marginal or no improvement over non-entangled inputs. Furthermore the P_e improvement, when it does occur, appears to decrease as the number of parallel channels increases.

Taken together, the conclusion to be drawn from our limited numerical study is that entanglement does not enhance the transmission of classical information over quantum channels in the same manner as entanglement aids quantum computation. The fact that unentangled signaling states appear to optimize the HSW channel capacity of parallel, independent quantum channels has important practical implications. Since product states are easier to create in the laboratory than input states with entanglement, one can potentially construct the optimal signaling states needed to achieve HSW channel capacity more easily than initially envisioned.

III Appendix A - Channel Descriptions

Recall that the Kraus representation is given by the set of Kraus matrices $\mathcal{A} = \{ A_i \}$ which represent the channel dynamics via the relation $\mathcal{E}(\rho) = \sum_i A_i \rho A_i^\dagger$. The normalization requirement for the Kraus matrices is $\sum_i A_i^\dagger \rho A_i = \mathcal{I}$. A channel is unital if it maps the identity to the identity. This requirement becomes, upon setting $\rho = \mathcal{I}$:

$$\sum_i A_i \rho A_i^\dagger = \sum_i A_i A_i^\dagger = \mathcal{I}.$$

For qubit channels, the set of Kraus operators, $\{ \mathcal{A}_i \}$ can mapped to a set of King-Ruskai-Szarek-Werner ellipsoid channel parameters $\{ t_k , \lambda_k \}$, where $k = 1, 2, 3$. The Kraus matrices and the King-Ruskai-Szarek-Werner ellipsoid channel parameters $\{ t_k , \lambda_k \}$ for the Two Pauli, Depolarization, and Amplitude Damping channels can be found in Appendix C of Chapter 2.

One Qubit Channel Kraus Representation

$$\mathcal{E}(\rho) = \sum_{i=1}^{i=3} A_i \rho A_i^\dagger.$$

The tensor product model of multiple channels leads to the following mathematical description in the Kraus framework.

Two Qubit Channel Kraus Representation

$$\mathcal{E}(\rho) = \sum_{i=1}^{i=3} \sum_{j=1}^{j=3} (A_i \otimes A_j) \rho (A_i \otimes A_j)^\dagger .$$

Under the independent channel action picture, the Kraus matrices for two uses of the channel are tensor products of the Kraus matrices for a single use of the channel.

Three Qubit Channel Kraus Representation

$$\mathcal{E}(\rho) = \sum_{i=1}^{i=3} \sum_{j=1}^{j=3} \sum_{k=1}^{k=3} (A_i \otimes A_j \otimes A_k) \rho (A_i \otimes A_j \otimes A_k)^\dagger .$$

IV Appendix B - Probability of Error Calculation

If we have two states with density matrices ρ_0 and ρ_1 and a priori probabilities p_0 and p_1 respectively, then the optimum set of POVMs can distinguish these two states with probability of error:

$$P_e \;=\; \frac{1}{2} \;-\; \frac{1}{4} \parallel p_1 \, \rho_1 \;-\; p_0 \, \rho_0 \parallel .$$

Here the notation $\parallel \mathcal{R} \parallel$ means that we take the sum of the absolute values of the eigenvalues of $\mathcal{R}$. That is, if the $\{ \lambda_i \}$ are the eigenvalues of $\mathcal{R}$, then

$$\parallel \mathcal{R} \parallel \;=\; \sum_i \; \mid \lambda_i \mid .$$

Note that $\mathcal{R}$ is a Hermitian matrix, so the eigenvalues are real. However, $\mathcal{R}$ is not a positive semi-definite matrix since $\mathcal{R}$ is the difference of two density matrices. Thus $\mathcal{R}$ can have negative eigenvalues.

Note that $P_e = 0$ iff the ρ_i have disjoint support, and $P_e = \frac{1}{2}$ iff $\rho_1 \;\equiv\; \rho_0$. For the detailed derivation of the probability of error formula, and the optimum POVMs, please see [44].

Chapter 4

Additivity of Holevo χ and HSW Channel Capacity

I Abstract

In this chapter, we consider fundamental properties of Holevo χ, maximizing functions over convex sets of quantum states, and quantum channel dynamics, and their possible role in the strict additivity of HSW channel capacity.

II Additivity of Holevo χ

A question one can ask is whether the Holevo χ quantity is additive across all bipartite ensembles. That is, given an ensemble $\mathcal{M} = \left\{p_i\,,\,\rho_{AB}^{(i)}\right\}$, where the $\rho_{AB}^{(i)}$ are possibly entangled states in a Hilbert space $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$, is $\chi_{AB} = \chi_A + \chi_B$?

In the two parallel channel scenario discussed in Chapter 1 and Chapter 3, we considered the channel output state as a bipartite system and found strict additivity of the Max Holevo χ quantity for two independent uses of certain classes of quantum channels. Below we show by example that for general bipartite ensembles, χ is *not* strictly additive. We give example ensembles $\mathcal{M} = \left\{\,p_i\,,\,\rho_{AB}^{(i)}\,\right\}$ where χ is subadditive, meaning $\chi_{AB} < \chi_A + \chi_B$ and super-additive, meaning $\chi_{AB} > \chi_A + \chi_B$.

The fact that χ is not by and in itself additive implies that strict $\mathcal{C}_1$ channel additivity derives either from the maximization operation or from some special property present in

channel output ensembles which is not present in more general bipartite ensembles, or perhaps some combination of these two conditions. This section was conceived to discuss what is, and is not, important in determining the issue of strict HSW channel additivity.

II.a Example Ensembles

Let $\mathcal{M} = \left\{ p_i , \rho_{AB}^{(i)} \right\}$ be the ensemble of bipartite states for which we seek to determine Holevo χ additivity. The specific question we address is the bipartite additivity, be it sub-additive, super-additive, or strict equality. Below we calculate χ_{AB}, χ_A, and χ_B and show by explicit construction that there exist ensembles for which $\chi_{AB} > \chi_A + \chi_B$ and $\chi_{AB} < \chi_A + \chi_B$. Here the ensemble $\mathcal{M}_A$ is the trace over subsystem B of $\mathcal{M}_{AB}$. Similarly for $\mathcal{M}_B$.

We begin by presenting some notation and a few definitions. Our ensembles are denoted $\mathcal{M} = \left\{ p_i , \rho_{AB}^{(i)} \right\}$. The relation between the $\rho_{AB}^{(i)}$ and ρ_{AB} is $\rho_{AB} = \sum_i p_i \rho_{AB}^{(i)}$. The general formula for the Holevo χ is $\chi = S(\rho_{AB}) - \sum_i p_i S\left(\rho_{AB}^{(i)}\right)$. We shall be interested in bipartite systems A and B, with each subsystem consisting of a single qubit. These considerations lead to the following quantities:

$$\chi_{AB} = S(\rho_{AB}) - \sum_i p_i S\left(\rho_{AB}^{(i)}\right)$$

$$\chi_A = S(\rho_A) - \sum_i p_i S\left(\rho_A^{(i)}\right)$$

$$\chi_B = S(\rho_B) - \sum_i p_i S\left(\rho_B^{(i)}\right)$$

where $\rho_A = Tr_B[\rho_{AB}]$ and $\rho_B = Tr_A[\rho_{AB}]$. Here S(-) denotes the von Neumann Entropy. If λ_i denote the eigenvalues of the density matrix ρ, then $S(\rho) = -\sum_i \lambda_i \log(\lambda_i)$.

II.b Example I - Holevo χ Super-Additivity

In this section, we give an example of a bipartite two-qubit ensemble $\mathcal{M}_{AB}$ which exhibits strict Holevo *super*-additivity, meaning $\chi_{AB} > \chi_A + \chi_B$. The number of elements in the ensemble $\mathcal{M}$ is 2.

$$\psi_{AB}^{(0)} = \frac{00 + 11}{\sqrt{2}} \qquad \rho_{AB}^{(0)} = |\psi_{AB}^{(0)}\rangle\langle\psi_{AB}^{(0)}|.$$

$$\psi_{AB}^{(1)} = \frac{01 + 10}{\sqrt{2}} \qquad \rho_{AB}^{(1)} = |\psi_{AB}^{(1)}\rangle\langle\psi_{AB}^{(1)}|.$$

$$p_0 = p_1 = \frac{1}{2}.$$

$$\rho_{AB}^{(0)} = \frac{1}{2}\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}.$$

$$\rho_{AB}^{(1)} = \frac{1}{2}\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

$$\rho_{AB} = \frac{1}{2}\rho_{AB}^{(0)} + \frac{1}{2}\rho_{AB}^{(1)} = \frac{1}{4}\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}.$$

$$\rho_A^{(0)} = Tr_B\left[\rho_{AB}^{(0)}\right] = \frac{1}{2}I_2 \qquad where \qquad I_2 = \frac{1}{2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

$$\rho_A^{(1)} = Tr_B\left[\rho_{AB}^{(1)}\right] = \frac{1}{2}I_2.$$

$$\rho_B^{(0)} = Tr_A\left[\rho_{AB}^{(0)}\right] = \frac{1}{2}I_2.$$

$$\rho_B^{(1)} = Tr_A\left[\rho_{AB}^{(1)}\right] = \frac{1}{2}I_2.$$

$$\rho_A \;=\; Tr_B[\rho_{AB}] \;=\; Tr_B\left[\,\frac{1}{2}\,\rho_{AB}^{(0)} \;+\; \frac{1}{2}\,\rho_{AB}^{(1)}\,\right] \;=\; \frac{1}{2}\,Tr_B\left[\,\rho_{AB}^{(0)}\,\right] \;+\; \frac{1}{2}\,Tr_B\left[\,\rho_{AB}^{(1)}\,\right]$$

$$=\; \frac{1}{2}\,\rho_A^{(0)} \;+\; \frac{1}{2}\,\rho_A^{(1)} \;=\; \frac{1}{4}\,I_2 \;+\; \frac{1}{4}\,I_2 \;=\; \frac{1}{2}\,I_2.$$

$$\rho_B \;=\; Tr_A[\rho_{AB}] \;=\; Tr_A\left[\,\frac{1}{2}\,\rho_{AB}^{(0)} \;+\; \frac{1}{2}\,\rho_{AB}^{(1)}\,\right] \;=\; \frac{1}{2}\,Tr_A\left[\rho_{AB}^{(0)}\right] \;+\; \frac{1}{2}\,Tr_A\left[\rho_{AB}^{(1)}\right]$$

$$=\; \frac{1}{2}\,\rho_B^{(0)} \;+\; \frac{1}{2}\,\rho_B^{(1)} \;=\; \frac{1}{4}\,I_2 \;+\; \frac{1}{4}\,I_2 \;=\; \frac{1}{2}\,I_2.$$

$$S_A^{(0)} \;=\; S\left(\rho_A^{(0)}\right) \;=\; 1 \qquad and \qquad S_A^{(1)} \;=\; S\left(\rho_A^{(1)}\right) \;=\; 1.$$

$$S_B^{(0)} \;=\; S\left(\rho_B^{(0)}\right) \;=\; 1 \qquad and \qquad S_B^{(1)} \;=\; S\left(\rho_B^{(1)}\right) \;=\; 1.$$

$$S_A \;=\; S(\rho_A) \;=\; 1 \qquad and \qquad S_B \;=\; S(\rho_B) \;=\; 1.$$

$$S_{AB}^{(0)} \;=\; S\left(\rho_{AB}^{(0)}\right) \;=\; 0. \qquad (Pure\ State)$$

$$S_{AB}^{(1)} \;=\; S\left(\rho_{AB}^{(1)}\right) \;=\; 0. \qquad (Pure\ State)$$

$$S_{AB} \;=\; S(\rho_{AB}) \;=\; 1.$$

$$\chi_{AB} \;=\; S_{AB} \;-\; \frac{1}{2}\,\left(\,S_{AB}^{(0)} \;+\; S_{AB}^{(1)}\,\right) \;=\; 1 \;-\; \frac{1}{2}\,(\,0 \;+\; 0\,) \;=\; 1.$$

$$\chi_A \;=\; S_A \;-\; \frac{1}{2}\,\left(\,S_A^{(0)} \;+\; S_A^{(1)}\,\right) \;=\; 1 \;-\; \frac{1}{2}\,(\,1 \;+\; 1\,) \;=\; 0.$$

$$\chi_B \;=\; S_B \;-\; \frac{1}{2}\,\left(\,S_B^{(0)} \;+\; S_{AB^{(1)}}\,\right) \;=\; 1 \;-\; \frac{1}{2}\,(\,1 \;+\; 1\,) \;=\; 0.$$

$$1 \;=\; \chi_{AB} \;>\; \chi_A \;+\; \chi_B \;=\; 0 \;+\; 0 \;=0.$$

$$\chi_{AB} \;>\; \chi_A \;+\; \chi_B\,.$$

(Below we define the Φ quantities. We list them here for future reference.)

$$\Phi_{AB} \;=\; S_A \;+\; S_B \;-\; S_{AB} \;=\; 1 \;+\; 1 \;-\; 1 \;=\; 1.$$

$$\Phi_{AB}^{(i)} \;=\; S_A^{(i)} \;+\; S_B^{(i)} \;-\; S_{AB}^{(i)} \;=\; 1 \;+\; 1 \;-\; 0 \;=\; 2.$$

For this ensemble, the Holevo χ quantity is *super-additive*.

II.c Example II - Holevo χ Sub-Additivity

In this section, we give an example of a bipartite two-qubit ensemble $\mathcal{M}_{AB}$ which exhibits strict Holevo *sub*-additivity, meaning $\chi_{AB} < \chi_A + \chi_B$. The number of elements in the ensemble $\mathcal{M}$ is 2.

$$\psi_{AB}^{(0)} = 00 \qquad \rho_{AB}^{(0)} = |\,\psi_{AB}^{(0)}\,\rangle\langle\,\psi_{AB}^{(0)}|.$$

$$\psi_{AB}^{(1)} = 11 \qquad \rho_{AB}^{(1)} = |\,\psi_{AB}^{(1)}\,\rangle\langle\,\psi_{AB}^{(1)}|.$$

$$p_0 = p_1 = \frac{1}{2}.$$

$$\rho_{AB}^{(0)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

$$\rho_{AB}^{(1)} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

$$\rho_{AB} = \frac{1}{2}\,\rho_{AB}^{(0)} + \frac{1}{2}\,\rho_{AB}^{(1)} = \frac{1}{2}\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

$$\rho_A^{(0)} = Tr_B\left[\rho_{AB}^{(0)}\right] = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

$$\rho_A^{(1)} = Tr_B\left[\rho_{AB}^{(1)}\right] = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

$$\rho_B^{(0)} = Tr_A\left[\rho_{AB}^{(0)}\right] = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

$$\rho_B^{(1)} = Tr_A\left[\rho_{AB}^{(1)}\right] = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

$$\rho_A \;=\; Tr_B[\rho_{AB}] \;=\; Tr_B\left[\,\frac{1}{2}\,\rho_{AB}^{(0)} \;+\; \frac{1}{2}\,\rho_{AB}^{(1)}\,\right] \;=\; \frac{1}{2}\,Tr_B\left[\rho_{AB}^{(0)}\right] \;+\; \frac{1}{2}\,Tr_B\left[\rho_{AB}^{(1)}\right]$$

$$=\; \frac{1}{2}\,\rho_A^{(0)} \;+\; \frac{1}{2}\,\rho_A^{(1)} \;=\; \frac{1}{2}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}.$$

$$\rho_B \;=\; Tr_A[\rho_{AB}] \;=\; Tr_A\left[\,\frac{1}{2}\,\rho_{AB}^{(0)} \;+\; \frac{1}{2}\,\rho_{AB}^{(1)}\,\right] \;=\; \frac{1}{2}\,Tr_A\left[\rho_{AB}^{(0)}\right] \;+\; \frac{1}{2}\,Tr_A\left[\rho_{AB}^{(1)}\right]$$

$$=\; \frac{1}{2}\,\rho_B^{(0)} \;+\; \frac{1}{2}\,\rho_B^{(1)} \;=\; \frac{1}{2}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}.$$

$$S_A^{(0)} \;=\; S\left(\rho_A^{(0)}\right) \;=\; 0 \qquad (Pure\ State) \qquad and \qquad S_A^{(1)} \;=\; S\left(\rho_A^{(1)}\right) \;=\; 0 \qquad (Pure\ State).$$

$$S_B^{(0)} \;=\; S\left(\rho_B^{(0)}\right) \;=\; 0 \qquad (Pure\ State) \qquad and \qquad S_B^{(1)} \;=\; S\left(\rho_B^{(1)}\right) \;=\; 0 \qquad (Pure\ State).$$

$$S_A \;=\; S(\rho_A) \;=\; 1 \qquad and \qquad S_B \;=\; S(\rho_B) \;=\; 1.$$

$$S_{AB}^{(0)} \;=\; S\left(\rho_{AB}^{(0)}\right) \;=\; 0. \qquad (Pure\ State)$$

$$S_{AB}^{(1)} \;=\; S\left(\rho_{AB}^{(1)}\right) \;=\; 0. \qquad (Pure\ State)$$

$$S_{AB} \;=\; S(\rho_{AB}) \;=\; 1.$$

$$\chi_{AB} \;=\; S_{AB} \;-\; \frac{1}{2}\,\left(\,S_{AB}^{(0)} \;+\; S_{AB}^{(1)}\,\right) \;=\; 1 \;-\; \frac{1}{2}\,(\,0 \;+\; 0\,) \;=\; 1.$$

$$\chi_A \;=\; S_A \;-\; \frac{1}{2}\,\left(\,S_A^{(0)} \;+\; S_A^{(1)}\,\right) \;=\; 1 \;-\; \frac{1}{2}\,(\,0 \;+\; 0\,) \;=\; 1.$$

$$\chi_B \;=\; S_B \;-\; \frac{1}{2}\,\left(\,S_B^{(0)} \;+\; S_{AB^{(1)}}\,\right) \;=\; 1 \;-\; \frac{1}{2}\,(\,0 \;+\; 0\,) \;=\; 1.$$

$$1 \;=\; \chi_{AB} \;<\; \chi_A \;+\; \chi_B \;=\; 1 \;+\; 1 \;=2.$$

$$\chi_{AB} \;<\; \chi_A \;+\; \chi_B.$$

$$\Phi_{AB} \;=\; S_A \;+\; S_B \;-\; S_{AB} \;=\; 1 \;+\; 1 \;-\; 1 \;=\; 1.$$

$$\Phi_{AB}^{(i)} \;=\; S_A^{(i)} \;+\; S_B^{(i)} \;-\; S_{AB}^{(i)} \;=\; 0 \;+\; 0 \;-\; 0 \;=\; 0.$$

For this ensemble, the Holevo χ quantity is *sub-additive*.

II.d Comments on Examples I and II

The heuristic reasoning behind these examples was the following. We were curious how bipartite entanglement influences the additivity of χ. Consider δ_χ defined below.

$$\delta_\chi \;=\; \chi_A \;+\; \chi_B \;-\; \chi_{AB}$$

$$= \left[\; S(\rho_A) \;-\; \sum_i p_i\, S\left(\rho_A^{(i)}\right) \right] + \left[\; S(\rho_B) \;-\; \sum_i p_i\, S\left(\rho_B^{(i)}\right) \right] - \left[\; S(\rho_{AB}) \;-\; \sum_i p_i\, S\left(\rho_{AB}^{(i)}\right) \right]$$

$$= S(\rho_A) \;+\; S(\rho_B) \;-\; S(\rho_{AB}) \;-\; \sum_i p_i \left[S\left(\rho_A^{(i)}\right) \;+\; S\left(\rho_B^{(i)}\right) \;-\; S\left(\rho_{AB}^{(i)}\right) \right]$$

$$= \Phi_{AB} \;-\; \sum_i p_i \left[\Phi_{AB}^{(i)} \right] \;=\; \Phi_{AB} \;-\; \left\langle\; \Phi_{AB}^{(i)} \;\right\rangle,$$

where we have defined the following quantities:

$$\Phi_{AB} \;=\; S(\rho_A) \;+\; S(\rho_B) \;-\; S(\rho_{AB})$$

$$\Phi_{AB}^{(i)} \;=\; S\left(\rho_A^{(i)}\right) \;+\; S\left(\rho_B^{(i)}\right) \;-\; S\left(\rho_{AB}^{(i)}\right)$$

$$\left\langle\; \Phi_{AB}^{(i)} \;\right\rangle \;=\; \sum_{i=0}^{i=k} p_i \left[S\left(\rho_A^{(i)}\right) \;+\; S\left(\rho_B^{(i)}\right) \;-\; S\left(\rho_{AB}^{(i)}\right) \right] \;=\; \sum_{i=0}^{i=k} p_i\, \Phi_{AB}^{(i)}$$

$$\delta\chi \;=\; \chi_A \;+\; \chi_B \;-\; \chi_{AB} \;=\; \Phi_{AB} \;-\; \left\langle\; \Phi_{AB}^{(i)} \;\right\rangle.$$

If system AB is a pure state, then system A and system B are entangled if and only if $S_A \;+\; S_B \;-\; S_{AB} \;>\; 0$. This is because for AB a pure state, $S_{AB} \;=\; 0$, and we always have $S_A \;\geq\; 0$, and $S_B \;\geq\; 0$. If A and B turn out to be pure states, then AB was a product state, AB was not entangled, and $S_A \;=\; 0$ and $S_B \;=\; 0$. In this case, $S_A \;+\; S_B \;-\; S_{AB} \;=\; 0$. However, if AB was entangled, then A and B will be mixed states, and $S_A \;>\; 0$ and $S_B \;>\; 0$, yielding $S_A \;+\; S_B \;-\; S_{AB} \;>\; 0$.

Looking to $\delta\chi$, we see χ is sub-additive ($\chi_A \;+\; \chi_B \;>\; \chi_{AB}$) if $\delta\chi \;>\; 0$.

$$Sub - Additivity \qquad \delta\chi \;>\; 0 \qquad \Phi_{AB} \;>\; \left\langle\; \Phi_{AB}^{(i)} \;\right\rangle.$$

$$Super - Additivity \qquad \delta\chi \; < \; 0 \qquad \Phi_{AB} \; < \; \left\langle \; \Phi_{AB}^{(i)} \; \right\rangle .$$

To obtain a $\delta\chi \; < \; 0$ in Example I, we have $\Phi_{AB} \; = \; S_A \; + \; S_B \; - \; S_{AB} \; = \; 1 \; + \; 1 \; - \; 1 \; = \; 1$ and $\Phi_{AB}^{(i)} \; = \; S_A^{(i)} \; + \; S_B^{(i)} \; - \; S_{AB}^{(i)} \; = \; 1 \; + \; 1 \; - \; 0 \; = \; 2$.

To demonstrate sub-additivity, we sought to use highly entangled pure states $\rho_{AB}^{(i)}$, which taken together form, *on average*, a "less entangled" (mixed) state ρ_{AB}. Hence the choice of Bell pairs in Example I.

For super-additivity of χ, we seek to have $\delta\chi \; < \; 0$. Here we sought to use *unentangled* pure states $\rho_{AB}^{(i)}$, which taken together form, on average, an "entangled" (mixed) state ρ_{AB}. Hence the choice of product states in Example II, which have $\Phi_{AB}^{(i)} \; = \; 0$. When the $\rho_{AB}^{(i)}$ are combined using the weighted probability distribution $\{ \; p_i \; \}$, we obtain a mixed state ρ_{AB}, with $\Phi_{AB} \; > \; 0$. Thus yielding $\Phi_{AB} \; > \; \left\langle \; \Phi_{AB}^{(i)} \; \right\rangle \; = \; 0$.

It is interesting to note that the quantity $\Phi_{AB} \; = \; S_A \; + \; S_B \; - \; S_{AB}$ is sometimes called the *information gain*. That is, it is the classical information gained by system A about the state of system B when system B is measured[43].

The implication to be drawn from these two examples, when combined with our strict additivity numerical results above, is that it is the *maximization* of χ across all possible channel output states that leads to the strict $\mathcal{C}_1$ additivity. That is, strict $\mathcal{C}_1$ additivity is not a natural property of χ, but results from the action of the channel restricting the set of possible output states available for the maximization process to work with.

II.e Asymptotic Example

One question that arises is whether these additivity results survive asymptotically. That is, if χ represents the Holevo quantity for a joint system of N subsystems, whether

$$\chi \; \gtrless \; \chi_1 \; + \; \chi_2 \; + \; \chi_3 \; + \; \cdots \; + \; \chi_N ,$$

or equality is approached as N $\rightarrow \infty$.

In this regard, Example II immediately generalizes. Take the number of ensemble members to be 2, and still consider single qubit subsystems, only now we shall use states with N qubits total. Let

$$\psi^{(0)} = 00\cdots00\,. \qquad (\textit{the 0 state N times})$$

$$\psi^{(1)} = 11\cdots11\,. \qquad (\textit{the 1 state N times})$$

The N party equivalent of the subsystem entropy, $S_A^{(i)}$, are zero, while the N-party equivalent of the average subsystem, ρ_A, has von Neumann entropy $S_A = 1$. This tells us that $\chi_A = S(\rho_A) - \sum_i p_i S\left(\rho_A^{(i)}\right) = 1$. This leads to:

$$\chi_1 + \chi_2 + \chi_3 + \cdots + \chi_N = N.$$

Since the states $\psi^{(i)}$ are pure states, the N party equivalent of the $S_{AB}^{(i)}$ are zero. Similarly, the N party equivalent of the average density matrix ρ_{AB} has von Neumann entropy $S_{AB} = 1$. Thus, the N party equivalent of χ_{AB}, which we have called χ with no subscripts, becomes $\chi = 1$. Putting it all together, we have the result below.

$$\chi = 1 < N = \chi_1 + \chi_2 + \chi_3 + \cdots + \chi_N.$$

Thus in this case, we have an N party sub-additive situation, even in the large N limit.

II.f Example III - *Max* Holevo χ

Consider taking the maximum of χ_{AB} across all possible two qubit ensembles $\left\{p_i\,,\,\rho_{AB}^{(i)}\right\}$. If you like, think of the maximum as corresponding to the determination of the $\mathcal{C}_1$ channel capacity of the tensor product of two single qubit identity channels $\mathcal{E}_A$ and $\mathcal{E}_B$, where $\mathcal{E}_A(\rho_A) = \rho_A$ and $\mathcal{E}_B(\rho_B) = \rho_B$. The HSW channel capacity $\mathcal{C}_1^{A\otimes B}$ is 2, and is achieved with the use of qubit pure states $\left\{\rho_{AB}^{(i)}\right\}$ which average to the two qubit state $\frac{1}{4}\mathcal{I}_4$. Furthermore, these pure states can be chosen to be equiprobable product states, for example

$\left\{p_1 = p_2 = p_3 = p_4 = \frac{1}{4},\, \rho_{AB}^{(i)}\right\}$, where the $\left\{\rho_{AB}^{(i)}\right\}$ correspond to the four possible product state spin up/down combinations of the AB subsystems:

$$\left\{\rho_{AB}^{(1)} \sim \uparrow_A\uparrow_B\,,\ \rho_{AB}^{(2)} \sim \uparrow_A\downarrow_B\,,\ \rho_{AB}^{(3)} \sim \downarrow_A\uparrow_B\,,\ \rho_{AB}^{(4)} \sim \downarrow_A\downarrow_B\right\}.$$

This ensemble choice yields A and B subchannel HSW capacities of $\mathcal{C}_1^A = \mathcal{C}_1^B = 1$. Thus strict additivity of channel capacity is achieved: $\mathcal{C}_1^{A\otimes B} = \mathcal{C}_1^A + \mathcal{C}_1^B$. The maximization of the two qubit channel output Holevo χ_{AB} quantity, together with the tensor product nature of the channel, yields a strictly additive HSW channel capacity.

II.g Example II Revisited

Consider the states used in Example II, and the same tensor product of two single qubit identity channels. Using the same equiprobable signalling states given in Example II, $\psi_{AB}^{(0)} = 00\ (\downarrow_A\downarrow_B)$, and $\psi_{AB}^{(1)} = 11\ (\uparrow_A\uparrow_B)$, *but with no Holevo χ maximization at the channel output*, we find $\chi_{AB} \neq \chi_A + \chi_B$. Additivity of Holevo χ, for this ensemble, without the maximization operation is *not* additive.

II.h Example I Revisited

Consider Example I with the following quantum channel.

$$\mathcal{E}_{AB}\left(\varphi_{AB}\right) = Tr\left[\varphi_{AB}\,\rho_{AB}^{(0)}\right]\,\rho_{AB}^{(0)} + Tr\left[\varphi_{AB}\,\rho_{AB}^{(1)}\right]\,\rho_{AB}^{(1)},$$

where the states $\rho_{AB}^{(0)}$ and $\rho_{AB}^{(1)}$ are from Example I. The HSW capacity of $\mathcal{C}_1^{AB}$ is 1. Note that $\mathcal{E}_{AB}$ cannot be written as the tensor product of two individual channels, $\mathcal{E}_{AB} \neq \mathcal{E}_A \otimes \mathcal{E}_B$. This means the errors induced on the two subchannels, $\mathcal{E}_A$ and $\mathcal{E}_B$, are *not* independent.

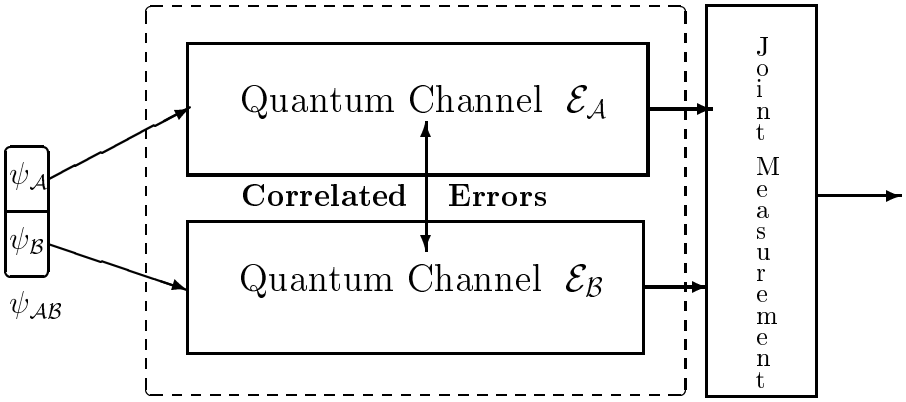

Figure 4.1: Two correlated channels.

Consider the partial trace channels $\mathcal{E}_A$ and $\mathcal{E}_B$ as defined below.

$$\mathcal{E}_A(\varphi_A) \;=\; Tr\left[\varphi_A \; Tr_B\left[\rho_{AB}^{(0)}\right]\right] \; Tr_B\left[\rho_{AB}^{(0)}\right] \;+\; Tr\left[\varphi_A \; Tr_B\left[\rho_{AB}^{(1)}\right]\right] \; Tr_B\left[\rho_{AB}^{(1)}\right]$$

and

$$\mathcal{E}_B(\varphi_B) \;=\; Tr\left[\varphi_B \; Tr_A\left[\rho_{AB}^{(0)}\right]\right] \; Tr_A\left[\rho_{AB}^{(0)}\right] \;+\; Tr\left[\varphi_A \; Tr_A\left[\rho_{AB}^{(1)}\right]\right] \; Tr_A\left[\rho_{AB}^{(1)}\right].$$

Recall from Example I that

$$Tr_A\left[\rho_{AB}^{(0)}\right] \;=\; Tr_A\left[\rho_{AB}^{(1)}\right] \;=\; Tr_B\left[\rho_{AB}^{(0)}\right] \;=\; Tr_B\left[\rho_{AB}^{(1)}\right] \;=\; \frac{1}{2}\mathcal{I}_2.$$

Thus

$$\mathcal{E}_A(\varphi_A) \;=\; \frac{1}{2}Tr\left[\varphi_A\right] \; \frac{1}{2}\mathcal{I}_2 \;+\; Tr\left[\varphi_{AB}\right] \; \frac{1}{2}\mathcal{I}_2 \;=\; \frac{1}{2}\mathcal{I}_2 \quad \forall\, \varphi_A.$$

Similarly

$$\mathcal{E}_B(\varphi_B) \;=\; \frac{1}{2}\mathcal{I}_2 \quad \forall\, \varphi_B.$$

Thus the partial trace channels $\mathcal{E}_A$ and $\mathcal{E}_B$ are point channels, and have HSW capacities $\mathcal{C}_1^A \;=\; \mathcal{C}_1^B \;=\; 0$. This leads to the superadditive capacity relation

$$\mathcal{C}_1^{AB} \;=\; 1 \;>\; 0 \;=\; \mathcal{C}_1^A \;+\; \mathcal{C}_1^B.$$

The lack of a tensor product structure in assembling the channel $\mathcal{E}_{AB}$ from the partial trace channels $\mathcal{E}_A$ and $\mathcal{E}_B$ led in this case to a super-additive HSW channel capacity relation.

We contend that it is the maximization of the Holevo χ at the channel output, *together with* the tensor product structure of the channel acting on the Hilbert space $\mathcal{H}_{AB}$ that leads to strict additivity of the HSW channel capacity.

III Discussion

In this chapter, our series of examples suggests the following conjecture.

Conjecture: The HSW channel capacity of two independent (tensor product) channels is strictly additive due to the combined action of the maximization over output Holevo χ and the independent, tensor product nature of the parallel channel construction.

Let us for the moment change gears, and present a geometrical picture of how strict additivity of HSW channel capacity *could* come about. We shall then return to our conjecture.

Consider two parallel, independent channels as in Figure 2 of Chapter 3. From Chapter 1, Section VII, we know that

$$\mathcal{C}_1^{A\otimes B} \;=\; min_{\psi_{AB}}\; max_{\rho_{AB}}\; \mathcal{D}(\rho_{AB}||\psi_{AB}) \;=\; \mathcal{D}(\,\varrho_k \,\|\, \varphi\,).$$

The last equality holds $\forall \varrho_k$ in an optimal output ensemble $\{p_k\,,\ \varrho_k\}$, and with $\varphi \;=\; \sum p_k\, \varrho_k$. A geometric view of the relation is shown in Figure 4.2.

The oval, including the oval interior, represents the convex set of possible quantum states output by the channel. The sole factor determining $\mathcal{C}_1^{A\otimes B}$ is the chord depicted above, where "distance" is measured by the relative entropy function. Here, due to the monotonicity of the relative entropy, ρ_{AB} and the ϱ_k are always on the boundary, while $\varphi \;=\; \sum p_k\, \varrho_k$ is always in the interior.

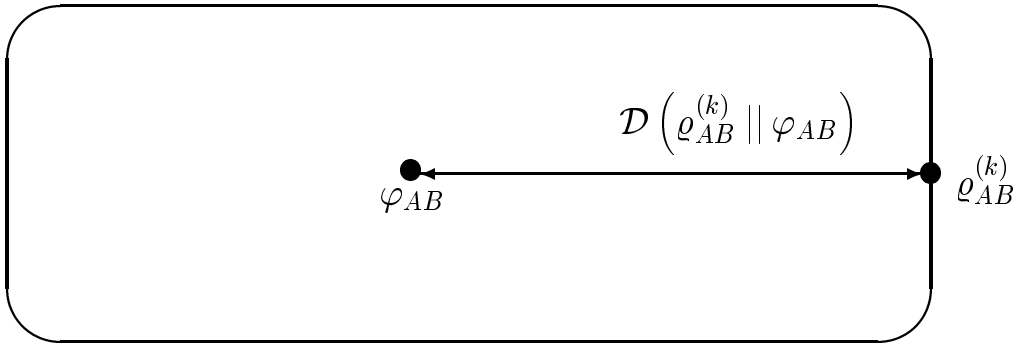

Figure 4.2: Geometric view of HSW channel capacity.

Consider what happens if we know there exists at least one optimal output signalling ensemble where the ensemble average output state φ_{AB} is a product state ($\varphi_{AB} = \varphi_A \otimes \varphi_B$), and at least one optimal output signalling state ϱ_{AB} in that same ensemble is a product state ($\varrho_{AB} = \varrho_A \otimes \varrho_B$). The HSW channel capacity is then strictly additive, since the Schumacher-Westmoreland Relative Entropy Lemmas from Section VII of Chapter 1 tell us the states ϱ_{AB} and φ_{AB} will satisfy the $Min\ Max$ criterion for $\mathcal{C}_1^{A\otimes B}$, namely

$$\mathcal{C}_1^{A\otimes B} = \mathrm{Min}_{\,\xi} \quad \mathrm{Max}_{\,\rho} \quad \mathcal{D}(\,\rho \,\|\, \xi\,) \quad = \quad \mathcal{D}(\,\varrho_{AB} \,\|\, \varphi_{AB}\,),$$

which together with the fact that the relative entropy function $\mathcal{D}(\cdots, \cdots)$ factors as shown in Equation I below, leads to the HSW channel capacity additivity result $\mathcal{C}_1^{A\otimes B} = \mathcal{C}_1^A + \mathcal{C}_1^B$.

$$\mathcal{C}_1^{A\otimes B} = Min_{\{\varrho_{AB}|\varrho_{AB}=\varrho_A\otimes\varrho_B\}}\, Max_{\{\varphi_{AB}|\varphi_{AB}=\varphi_A\otimes\varphi_B\}}\, \mathcal{D}\,(\varrho_{AB}||\varphi_{AB}) \qquad \text{(I)}$$

$$= Min_{\{\varrho_{AB}|\varrho_{AB}=\varrho_A\otimes\varrho_B\}}\, Max_{\{\varphi_{AB}|\varphi_{AB}=\varphi_A\otimes\varphi_B\}}\, Tr_{AB}\,[\,\varrho_{AB}\,\log(\,\varrho_{AB}\,) \;-\; \varrho_{AB}\,\log(\,\varphi_{AB}\,)]$$

$$= Min_{\{\varrho_A\}}\, Min_{\{\varrho_B\}}\, Max_{\{\varphi_A\}}\, Max_{\{\varphi_B\}}\, \ldots$$

$$Tr_{AB}\,[\,(\varrho_A\otimes\varrho_B)\,\log(\,\varrho_A\otimes\varrho_B\,) \;-\; (\varrho_A\otimes\varrho_B)\,\log(\,\varphi_A\otimes\varphi_B\,)]$$

$$= Min_{\{\varrho_A\}}\, Min_{\{\varrho_B\}}\, Max_{\{\varphi_A\}}\, Max_{\{\varphi_B\}}\, \ldots$$

$$(\,Tr_{AB}\,[\,(\varrho_A\otimes\varrho_B)\,(\log(\,\varrho_A\,)\otimes\mathcal{I}_B\,) \;+\; (\varrho_A\otimes\varrho_B)\,(\mathcal{I}_A\otimes\log(\,\varrho_B\,))\,]$$

$$- Tr_{AB}[(\varrho_A \otimes \varrho_B)(\log(\varphi_A) \otimes \mathcal{I}_B) + (\varrho_A \otimes \varrho_B)(\mathcal{I}_A \otimes \log(\varphi_B))])$$

$$= Min_{\{\varrho_A\}} Min_{\{\varrho_B\}} Max_{\{\varphi_A\}} Max_{\{\varphi_B\}} \ldots$$

$$(Tr_A[\varrho_A \log(\varrho_A)] Tr_B[\varrho_B \mathcal{I}_B] + Tr_A[\varrho_A \mathcal{I}_A] Tr_B[\varrho_B \log(\varrho_B)]$$

$$- Tr_A[\varrho_A \log(\varphi_A)] Tr_B[\varrho_B \mathcal{I}_B] - Tr_A[\varrho_A \mathcal{I}_A] Tr_B[\varrho_B \log(\varphi_B)])$$

$$= Min_{\{\varrho_A\}} Min_{\{\varrho_B\}} Max_{\{\varphi_A\}} Max_{\{\varphi_B\}} \ldots$$

$$(Tr_A[\varrho_A \log(\varrho_A)] - Tr_A[\varrho_A \log(\varphi_A)] + Tr_B[\varrho_B \log(\varrho_B)] - Tr_B[\varrho_B \log(\varphi_B)])$$

$$= Min_{\{\varrho_A\}} Min_{\{\varrho_B\}} Max_{\{\varphi_A\}} Max_{\{\varphi_B\}} \quad (\mathcal{D}(\varrho_A||\varphi_A) + \mathcal{D}(\varrho_B||\varphi_B))$$

$$= Min_{\{\varrho_A\}} Min_{\{\varrho_B\}} Max_{\{\varphi_A\}} Max_{\{\varphi_B\}} \mathcal{D}(\varrho_A||\varphi_A) \quad +$$

$$Min_{\{\varrho_A\}} Min_{\{\varrho_B\}} Max_{\{\varphi_A\}} Max_{\{\varphi_B\}} \mathcal{D}(\varrho_B||\varphi_B)$$

$$= Min_{\{\varrho_A\}} Max_{\{\varphi_A\}} \mathcal{D}(\varrho_A||\varphi_A) \quad + \quad Min_{\{\varrho_B\}} Max_{\{\varphi_B\}} \mathcal{D}(\varrho_B||\varphi_B) = \mathcal{C}_1^A + \mathcal{C}_1^B.$$

Furthermore the fact that $\mathcal{C}_1^{A\otimes B} = \mathcal{C}_1^A + \mathcal{C}_1^B$ trivially implies that there exists an optimal ensemble consisting entirely of product signalling states $\varrho_{AB}^{(i)} = \varrho_A^{(i)} \otimes \varrho_B^{(i)}$, namely the ensemble constructed from the tensor products of the individual channel optimum output signalling states $\left\{ \varrho_A^{(j)} \right\}$ and $\left\{ \varrho_B^{(k)} \right\}$, and by the uniqueness of the optimal ensemble average output state, that the optimal ensemble average output state φ_{AB} is a product state ($\varphi_{AB} = \varphi_A \otimes \varphi_B$). Thus $\mathcal{C}_1^{A\otimes B} = \mathcal{C}_1^A + \mathcal{C}_1^B$ if and only if the optimal ensemble average output state φ_{AB} is a product state ($\varphi_{AB} = \varphi_A \otimes \varphi_B$) *and* there exists at least one optimal output signalling ensemble which contains at least one optimal signalling state ϱ_{AB} which is a product state ($\varrho_{AB} = \varrho_A \otimes \varrho_B$).

Our motivation for making the product state assumption for the optimal output signalling states $\left\{ \varrho_{AB}^{(i)} \right\}$ and the average output state of an optimal ensemble φ_{AB}, was inspired by the numerical work in Chapter 3. There we found the optimum output signalling states were always product states. In addition, our work with diagonal unital qudit channels indicated the optimum average output state φ_{AB}, was always $\frac{1}{d}\mathcal{I}_d$, a product state. However, we

were unable to *analytically* prove that, in general, there always exists an optimal output ensemble with at least one signalling state which is a product state, even for the restricted class of diagonal unital qudit channels discussed in Chapter 1.

For diagonal unital qudit channels, the existence of an optimal output signalling state ϱ_{AB} which is a product state ($\varrho_{AB} = \varrho_A \otimes \varrho_B$) is equivalent to the satisfaction of the minimum entropy theorem of Chapter 1, Section XIII. This is due to the fact that in this case we know that the average output ensemble state is a product state $\left(\varphi_{AB} = \frac{1}{d} \mathcal{I}_d \right)$. Thus we have the following three if and only if relationships.

1) $\exists$ an optimal signalling ensemble with an optimal output signalling state ϱ_{AB} which is a product state ($\varrho_{AB} = \varrho_A \otimes \varrho_B$)

$\overset{From\ Above}{\Longleftrightarrow}$ 2) $\mathcal{C}_1^{A \otimes B} = \mathcal{C}_1^A + \mathcal{C}_1^B$

$\overset{From\ Chapter\ 1,\ Section\ XIII}{\Longleftrightarrow}$ 3) $min_{\rho_{\mathcal{AB}}} \mathcal{S}\left(\mathcal{E}(\rho_{\mathcal{AB}})\right) = min_{\rho_{\mathcal{A}}} \mathcal{S}\left(\mathcal{E}(\rho_{\mathcal{A}})\right) + min_{\rho_{\mathcal{B}}} \mathcal{S}\left(\mathcal{E}(\rho_{\mathcal{B}})\right)$

In our super-additive HSW channel capacity example above (*Example I Revisited*), the correlated channels scenario allowed only entangled states on the boundary. With no boundary product states available as optimal output signalling states, we were forced to use entangled boundary signalling states in the optimal output ensemble $\left\{ \varrho_{AB}^{(k)} \right\}$ and in the relation $\mathcal{C}_1^{A \otimes B} = \mathcal{D}\left(\rho_{AB}^{(k)} || \varphi_{AB} \right)$. Entangled output signalling states $\left\{ \varrho_{AB}^{(k)} \right\}$ do not lead to the relative entropy factorization shown above in Equation I, thereby yielding an example of non-additive HSW channel capacity.

Returning to our conjecture above and the picture whereby we seek to maximize the chord shown in Figure 4.2, our limited numerical analyses in Chapter 3 lead us to conjecture that the maximization of the output ensemble Holevo χ in the case of tensor product channel constructions, drives the optimization process to optimal output signalling states on the boundary of the set of channel output states which are product states. One possible scenario is that the maximization of the output χ drives the optimization process determining the optimal output signalling states to extremal states on the boundary, and these extremal

states are always product states.

In summary, if the (unique) average output ensemble state is a product state, and in addition if one or more of the optimal output signalling states in *any* optimal output ensemble is a product state, than our factorization of the relative entropy function above indicates the HSW channel capacity of the tensor product of the two channels will be additive. This occurred in Example III, where we were able to find an optimal output signalling ensemble which consisted of product states, and had an average output ensemble state which was a product state. The HSW capacity was strictly additive, despite in this case the existence of optimal output ensembles which did not contain any optimal output signalling states which were product states. However, the general conditions on a channel $\mathcal{E}_{AB}$ which would ensure that the (unique) average output ensemble state φ_{AB} is a product state, *and* that one or more of the optimal output signalling states ϱ_{AB} in at least one optimal output ensemble is a product state, remains an open question.

Thank you for your time.

Bibliography

[1] John A. Cortese, "The Holevo-Schumacher-Westmoreland Channel Capacity for a Class of Qudit Unital Channels", 2002, LANL ArXiV e-print quant-ph/0211093.

[2] M. A. Nielsen, and Isaac L. Chuang, *Quantum Information and Computation*, Cambridge University Press, New York, 2000.

[3] M-D Choi, "Completely Positive Linear Maps on Complex Matrices", *Lin. Alg. Appl.* **10**, 285-290 (1975).

[4] M-D Choi, "A Schwarz Inequality for Positive Linear Maps on C^* Algebras", *Ill. J. Math.* **18**, 565-574 (1974).

[5] David J. Griffiths, *Introduction To Quantum Mechanics*, Prentice Hall Inc., 1995.

[6] Eugen Merzbacher, *Quantum Mechanics*, Third Edition, John Wiley and Sons, Inc., 1998.

[7] David Parks, *Introduction to the Quantum Theory*, Third Edition, McGraw Hill, 1992.

[8] Christopher Fuchs, "Quantum information: How much information in a state vector?", 1996, LANL ArXiV e-print quant-ph/9601025.

[9] Christopher Fuchs, " Information Gain vs. State Disturbance in Quantum Theory", 1996, LANL ArXiV e-print quant-ph/9605014.

[10] Christopher Fuchs, "Information Gain vs. State Disturbance in Quantum Theory", 1996, LANL ArXiV e-print quant-ph/9611010.

[11] Christopher Fuchs, "Information Tradeoff Relations for Finite-Strength Quantum Measurements", 2000, LANL ArXiV e-print quant-ph/0009101.

[12] B. W. Schumacher and M. Westmoreland, "Sending classical information via noisy quantum channels," *Physical Review A*, Volume **56**, Pages 131–138, 1997.

[13] A. S. Holevo, "The capacity of quantum channel with general signal states," *IEEE Transactions on Information Theory*, Volume **44**, Pages 269, 1998.

[14] P. Hausladen, R. Jozsa, B. W. Schumacher, M. Westmoreland, and W. K. Wootters, "Classical information capacity of a quantum channel," *Physical Review A*, Volume **54**, Pages 1869–1876, 1996.

[15] Thomas M. Cover and Joy A. Thomas, *Elements of Information Theory*, John Wiley and Sons, New York, 1991.

[16] B. W. Schumacher and M. Westmoreland, "Optimal Signal Ensembles", 1999, LANL ArXiV e-print quant-ph/9912122.

[17] A. S. Holevo, IEEE Trans. Information Theory, Volume **44**, Page 269, (1998).

[18] Peter Shor, "Equivalence of Additivity Questions in Quantum Information Theory" 2003, LANL ArXiV e-print quant-ph/0305035.

[19] Christopher King, "Additivity for a class of unital qubit channels", 2001, LANL ArXiV e-print quant-ph/0103156.

[20] Wilbur B. Davenport, Jr. and William L. Root, *An Introduction to the Theory of Random Signals and Noise*, McGraw-Hill, 1958.

[21] Alberto Leon-Garcia, *Probability and Random Processes for Electrical Engineering*, Second Edition, Addison-Wesley Publishing, 1994.

[22] B. W. Schumacher and M. Westmoreland, "Relative Entropy in Quantum Information Theory" 2000, LANL ArXiV e-print quant-ph/0004045.

[23] Masanori Ohya and Denes Petz, *Quantum Entropy and Its Use*, Springer-Verlag, 1993.

[24] Masanori Ohya, Denes Petz and N. Watanabe, Prob. Math. Stats., Volume **17**, Page 170, 1997.

[25] Christopher King and Mary Beth Ruskai, "Minimal Entropy of States Emerging from Noisy Quantum Channels", 1999, LANL ArXiV e-print quant-ph/9911079.

[26] Mary Beth Ruskai, Stanislaw Szarek, and Elisabeth Werner, "An Analysis of Completely-Positive Trace-Preserving Maps on 2 by 2 Matrices", 2001, LANL ArXiV e-print quant-ph/0101003.

[27] J. F. Cornwell, *Group Theory In Physics : An Introduction*, Academic Press, 1997.

[28] Chris J. Isham, *Modern Differential Geometry for Physicists*, World Scientific, Second Edition, 1999, Page 187.

[29] Paul R. Halmos, *Finite-Dimensional Vector Spaces*, Springer Verlag, 1987, Page 95-96.

[30] Roger A. Horn and Charles R. Johnson, *Matrix Analysis*, Cambridge University Press, 1986, Page 71.

[31] Dr. Eric Rains, Private Communication.

[32] A. S. Holevo, "Remarks on the classical capacity of quantum channel", 2002, LANL ArXiV e-print quant-ph/0212025.

[33] M. J. Donald, Mathematical Proceedings of the Cambridge Philosophical Society, Volume **101**, Page 363, 1987.

[34] John P. Preskill, *Course Notes for California Institute of Technology Physics 219*, Chapter 7, http://www.theory.caltech.edu/~preskill.

[35] Paul R. Halmos, *Finite-Dimensional Vector Spaces*, Springer Verlag, 1987, Page 124, Theorem 2.

[36] Roger A. Horn and Charles R. Johnson, *Matrix Analysis*, Cambridge University Press, 1986, Page 182.

[37] John Cortese, "Relative Entropy and Single Qubit Holevo-Schumacher-Westmoreland Channel Capacity", 2002, LANL ArXiV e-print quant-ph/0207128.

[38] Charles H. Bennett, Christopher A. Fuchs, John A. Smolin, "Entanglement-Enhanced Classical Communication on a Noisy Quantum Channel" 1996, LANL ArXiV e-print quant-ph/9611006.

[39] Christopher King and Mary Beth Ruskai, "Capacity of Quantum Channels Using Product Measurements", 2000, LANL ArXiV e-print quant-ph/0004062.

[40] Christopher King, Michael Nathanson, and Mary Beth Ruskai, "Qubit Channels Can Require More Than Two Inputs To Achieve Capacity", 2001, LANL ArXiV e-print quant-ph/0109079.

[41] Mary Beth Ruskai, "Qubit Entanglement Breaking Channels", 2003, LANL ArXiV e-print quant-ph/0302032.

[42] Michael Horodecki, Peter W. Shor, Mary Beth Ruskai, "General Entanglement Breaking Channels", 2003, LANL ArXiV e-print quant-ph/0302031.

[43] Vlatko Vedral, "Landauer's erasure, error correction, and entanglement," 1999, LANL ArXiV e-print quant-ph/9903049.

[44] Carl W. Helstrom *Quantum Detection and Estimation Theory*, Academic Press, New York, 1976.